Manual of Cardiovascular Proteomics

Giulio Agnetti
Merry L. Lindsey
D. Brian Foster

Editors

Manual of Cardiovascular Proteomics

Giulio Agnetti • Merry L. Lindsey
D. Brian Foster

Editors

Manual of Cardiovascular Proteomics

Editors
Giulio Agnetti
Division of Cardiology
Johns Hopkins University School of Medicine
Baltimore
Maryland
USA

Department of Biomedical and
Neuromotor Sciences (DIBINEM)
University of Bologna
Bologna
Italy

Merry L. Lindsey
Dept. of Physiology & Biophysics
University Mississippi Medical Center
Jackson
Mississippi
USA

D. Brian Foster
Division of Cardiology
Johns Hopkins University School of Medicine
Baltimore
Maryland
USA

To Giulio Stoppioni, Alicia Taylor, and all others who have died of cardiovascular disease. To our families who have supported our pursuit of a career in cardiovascular proteomics, and to our colleagues who have helped advance the field.

Preface

The proteome, originally defined as the protein complement of the genome, can be studied by a complex array of technologies commonly encapsulated under the term *proteomics*. Of these technologies, mass spectrometry has single handedly revolutionized our ability to explore biological systems and provides a major advancement for research efforts. The realization that protein post-translational modifications are heterogeneous, ubiquitous, and dynamic offers an expanded view to examine pathophysiological processes. This concept has compelling implications for the understanding of diseases, particularly those that are sporadic in nature and for which a genetic determinant is not readily identified. The low correlation between gene expression and phenotype in health and disease can be explained by changes at the protein post-transcriptional and post-translational levels. This is particularly relevant in the settings of cardiovascular disease. For instance, increasing evidence supports the idea that cardiovascular diseases, such as heart failure, are proteinopathies. Therefore, there is a clear need to evaluate proteins and their modifications, to better understand cardiovascular disease at the molecular level.

This book is a response to that clarion call. We highlight the remarkable advances that have contributed to the development of proteomics over the last two decades. Enormous leaps have accompanied the implementation of new instruments and analytical approaches. As a result, hurdles along the path from data acquisition to biological insight have been dramatically lowered. The aims of this manual are to make these new methods and technologies understandable to scientists who are new to the proteomics arena and to accelerate knowledge dissemination. This book covers the full measure of topics that define modern cardiovascular proteomics, including but not limited to experimental design, sample preparation and separation, mass spectrometry technologies, protein identification and quantitation, as well as statistical, pathway, and network analyses of proteomics results.

While many of the concepts, methods, and technologies presented here can be widely applied across fields of biomedical research, including cancer and brain research, this book has been tailored for the cardiovascular researcher; the many facets of cardiovascular disease are reflected in these chapters. All proteomes are not created equal, and cardiovascular proteomics presents a host of unique

challenges, not the least of which is the wide dynamic range of protein abundances present in the cardiac muscle and in plasma. This book addresses many of the technical subtleties to be considered when embarking on a cardiovascular proteomic study, which serves as a guide for scientists entering the field.

This is the dawn of a new century where technology serves biology in an unprecedented way, and the tools currently available in the toolbox provide a rapid means to impact cardiovascular research. The contributing authors are among the most established scientists in the field of cardiovascular proteomics and have experienced the proteomic revolution first hand. We are extremely grateful to all of the authors for dedicating their precious time to share their deep knowledge and help us compile this manual.

Finally, we are immensely grateful for the strong historical foundation on which these chapters are built, laid by scientists pursuing their passion before the term proteomics was even coined. Among them, we would like acknowledge Dr. Jennifer E. Van Eyk, Dr. Peipei Ping, and Dr. Michael J. Dunn for their rich scientific contributions, their enthusiastic mentoring of new generations, and their tireless dedication to the advancement of proteomics as an essential field of cardiovascular discovery. Today, the state of the field is stronger for their efforts, and its vitality is directly reflected in this book. We hope that you find it useful for your research.

Baltimore, MD, USA   Giulio Agnetti
Baltimore, MD, USA   D. Brian Foster
Jackson, MS, USA   Merry L. Lindsey

Contents

Contributors

Giulio Agnetti, PhD, FAHA Laboratory of Cardiovascular Biochemistry, Division of Cardiology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy

D. Kent Arrell, PhD Center for Regenerative Medicine and Marriott Heart Disease Research Program, Division of Cardiovascular Diseases, Department of Medicine, Department of Molecular Pharmacology and Experimental Therapeutics, and Department of Medical Genetics, Mayo Clinic, Rochester, MN, USA

Jennifer A. Beckman, MSN, BSN Division of Cardiology, Department of Internal Medicine, University of Washington Medical Center, Seattle, WA, USA

Francesca Brambilla, PhD Department of Proteomics and Metabolomics, Institute for Biomedical Technologies-National Research Council, Milan, Segrate, Italy

Emilio Camafeita, PhD Cardiovascular Proteomic Lab and Proteomic Unit, Vascular Pathophysiology Area, Centro Nacional De Investigaciones Cardiovasculares, Madrid, Spain

Jenice Cheah Department of Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, CA, USA

Heaseung Sophia Chung, MS Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Luis Miguel Blanco-Colio, PhD   Vascular Research Lab, FIIS-Fundacion Jimenez Diaz, Autonoma University of Madrid, Madrid, Spain

Jake Cosme, HBSc Department of Physiology, University of Toronto and Toronto General Hospital Research Institute, Toronto, ON, Canada

Gemma Currie, MB, ChB Institute of Cardiovascular and Medical Sciences,
BHF Glasgow Cardiovascular Research Centre, University of Glasgow,
Glasgow, UK

Lisandra E. de Castro Brás, PhD Department of Physiology,
The Brody School of Medicine, East Carolina University,
Greenville, NC, USA

Christian Delles, MD, FRCP, FAHA, FBHS Institute of Cardiovascular
and Medical Sciences, BHF Glasgow Cardiovascular Research Centre,
University of Glasgow, Glasgow, UK

Dario Di Silvestre, PhD Department of Proteomics and Metabolomics,
Institute for Biomedical Technologies-National Research Council,
Milan, Segrate, Italy

Michael J. Dunn, BSc, PhD Conway Institute, University College Dublin,
Dublin, Ireland

Stephen Farris, MD Division of Cardiology, Department of Internal Medicine,
University of Washington Medical Center, Seattle, WA, USA

Clifford Folmes, PhD Division of Cardiovascular Diseases,
Department of Medicine, Center for Regenerative Medicine, Mayo Clinic,
Rochester, MN, USA

D. Brian Foster, PhD Laboratory of Cardiovascular Biochemistry, Division of
Cardiology, Department of Medicine, The Johns Hopkins University School of
Medicine, Baltimore, MD, USA

Sarah Franklin, PhD Cardiology Division, Department of Internal Medicine,
The Nora Eccles Harrison Cardiovascular Research and Training Institute,
University of Utah, Salt Lake City, UT, USA

Department of Biochemistry, University of Utah, Salt Lake City, UT, USA

Ying Ge, PhD Cellular and Molecular Pathology Training Program,
University of Wisconsin-Madison, Madison, WI, USA

Human Proteomics Program, University of Wisconsin-Madison,
Madison, WI, USA

Jennifer E. Gilda, BS Department of Neurobiology, Physiology, and Behavior,
University of California, Davis, Davis, CA, USA

Aldrin V. Gomes, PhD Department of Neurobiology, Physiology, and Behavior,
University of California, Davis, Davis, CA, USA

Anthony Gramolini, PhD Department of Physiology, University of Toronto
and Toronto General Hospital Research Institute, Toronto, ON, Canada

Department of Physiology, Ted Rogers Centre for Heart Research,
University of Toronto and University Health Network, Toronto, ON, Canada

Zachery R. Gregorich, BS Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI, USA

Molecular Pharmacology Training Program, University of Wisconsin-Madison, Madison, WI, USA

Rebekah L. Gundry, PhD Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA

Ronald J. Holewinski, PhD The Advanced Clinical Biosystems Research Institute, Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Tuleagh Innes-Gawn Department of Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, CA, USA

Elena Burillo, PhD Vascular Research Lab, FIIS-Fundacion Jimenez Diaz, Autonoma University of Madrid, Madrid, Spain

Shanxi Jiang, BS Department of Anesthesiology, David Geffen School of Medicine and UCLA, Los Angeles, CA, USA

Kai Kammers, PhD Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Mark J. Kohr, PhD Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Erin M. Kropp Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA

Nicole M. Lane, BS, BA Cellular and Molecular Pathology Training Program, University of Wisconsin-Madison, Madison, WI, USA

Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI, USA

Merry L. Lindsey, PhD Department of Physiology and Biophysics, Mississippi Center for Heart Research, University of Mississippi Medical Center, Jackson, MS, USA

Research Service, G.V. (Sonny) Montgomery Veterans Affairs Medical Center, Jackson, MS, USA

Diego Martínez-López Vascular Research Lab, FIIS-Fundacion Jimenez Diaz, Autonoma University of Madrid, Madrid, Spain

Department of Nephrology, Vascular Pathology, Hypertension, and Diabetes, IIS-Fundación Jiménez Díaz-UAM, Madrid, Spain

Juan Antonio Lopez, PhD Cardiovascular Proteomic Lab and Proteomic Unit, Vascular Pathophysiology Area, Centro Nacional De Investigaciones Cardiovasculares, Madrid, Spain

Claudius Mahr, DO, FACC, FESC Division of Cardiology, Department of
Internal Medicine, University of Washington Medical Center, Seattle, WA, USA

Department of Internal Medicine, University of Washington, Seattle, WA, USA

Roxana Martinez-Pinna, PhD Vascular Research Lab, FIIS-Fundacion Jimenez
Diaz, Autonoma University of Madrid, Madrid, Spain

Peter Matt, MD, FETCS Aortic Research Group, Division of Cardiac Surgery,
University Hospital Basel, University of Basel, Basel, Switzerland

Pierluigi Mauri, PhD Department of Proteomics and Metabolomics,
Institute for Biomedical Technologies-National Research Council,
Milan, Segrate, Italy

Laboratory of Medical Science, Scuola Superiore Sant'Anna,
Institute of Life Sciences, Pisa, Italy

Mickey R. Miller, PhD Department of Cardiology, The Nora Eccles Harrison
Cardiovascular Research and Training Institute, University of Utah,
Salt Lake City, UT, USA

Anne M. Murphy, MD Division of Cardiology, Department of Pediatrics,
Johns Hopkins University School of Medicine, Baltimore, MD, USA

Christopher I. Murray, PhD Department of Anesthesiology, Pharmacology
and Therapeutics, University of British Columbia, Vancouver, BC, Canada

Life Sciences Centre, Vancouver, BC, Canada

Kyriakos N. Papanicolaou, PhD Division of Cardiology, Department
of Medicine, Johns Hopkins University, Johns Hopkins School of Medicine,
Baltimore, MD, USA

Laurie Parker, PhD Department of Biochemistry, Molecular Biology,
and Biophysics, University of Minnesota, Minneapolis, MN, USA

Sarah J. Parker, PhD The Advanced Clinical Biosystems Research Institute,
Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Genaro Pimienta, PhD, MSc Biomedical Innovation Department,
Center for Scientific Research and Higher Education at Ensenada, Ensenada,
Baja California, Mexico

Jun Qu, PhD Department of Biochemistry, University at Buffalo,
State University of New York, Buffalo, NY, USA

Department of Pharmaceutical Sciences, University at Buffalo,
State University of New York, Buffalo, NY, USA

State of New York Center for Excellence in Bioinformatics and Life Sciences,
Buffalo, NY, USA

Genaro Antonio Ramirez-Correa, MD, PhD Division of Cardiology, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Ingo Ruczinski, PhD Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Shichen Shen, BS Department of Biochemistry, University at Buffalo, State University of New York, Buffalo, NY, USA

State of New York Center for Excellence in Bioinformatics and Life Sciences, Buffalo, NY, USA

Xiaomeng Shen, BS Department of Biochemistry, University at Buffalo, State University of New York, Buffalo, NY, USA

State of New York Center for Excellence in Bioinformatics and Life Sciences, Buffalo, NY, USA

Stanley V. Smith, PhD Department of Pharmacology and Toxicology, University of Mississippi Medical Center, Jackson, MS, USA

April Stempien-Otero, MD Division of Cardiology, Department of Internal Medicine, University of Washington Medical Center, Seattle, WA, USA

Irina Tchernyshyov, MSc The Advanced Clinical Biosystems Research Institute, Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Andre Terzic, MD, PhD Center for Regenerative Medicine and Marriott Heart Disease Research Program, Division of Cardiovascular Diseases, Department of Medicine, Department of Molecular Pharmacology and Experimental Therapeutics, and Department of Medical Genetics, Mayo Clinic, Rochester, MN, USA

Jennifer E. Van Eyk, PhD, FAHA, FISHR The Advanced Clinical Biosystems Research Institute, Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Jesus Vazquez, PhD Cardiovascular Proteomic Lab and Proteomic Unit, Vascular Pathophysiology Area, Centro Nacional De Investigaciones Cardiovasculares, Madrid, Spain

Vidya Venkatraman, MS The Advanced Clinical Biosystems Research Institute, Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Jose Luis Martin-Ventura, PhD Vascular Research Lab, FIIS-Fundacion Jimenez Diaz, Autonoma University of Madrid, Madrid, Spain

Thomas M. Vondriska, PhD Department of Anesthesiology,
David Geffen School of Medicine and UCLA, Los Angeles, CA, USA

Matthew Waas Department of Biochemistry, Medical College of Wisconsin,
Milwaukee, WI, USA

Ian Wright, BSc, CBiol. MRSB Advanced Clinical Biosystems Research
Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Strategic Innovations LLC, Earleville, MD, USA

Mingguo Xu, MD, PhD Department of Pediatric Cardiology,
Shenzhen Children's Hospital, Shenzhen, Guangdong, China

Division of Cardiology, Department of Pediatrics, Johns Hopkins University
School of Medicine, Baltimore, MD, USA

Chapter 1
A Historical Perspective on Cardiovascular Proteomics

Giulio Agnetti and Michael J. Dunn

Abstract The "Manual of Cardiovascular Proteomics" is the result of the concerted effort of many experts in the field and it addresses the core technologies and approaches that have been implemented since its birth. Although each chapter can be read or studied independently of the others; depending on the level of interest, the whole manual should provide a detailed overview on what is available to the modern scientist who wants to embark on a cardiovascular proteomic expedition. Chapter 1 provides the historical perspective and describes the landmark discoveries that propelled the field forward, along with considerations on how to chose a specific approach and what the first steps to complete a proteomic experiment successfully should be. We hope that you will enjoy the first edition and are looking forward to your feedback in order to improve future editions.

Introduction: The Renaissance of Protein Biochemistry

At the end of the last century, in September 1994 to be exact, a new term was publicly announced during a small meeting titled "2D Electrophoresis: From Protein Maps to Genomes" held in the charming Tuscan hills (*Siena*, *Italy*). At this meeting, the term "Proteome" was defined by Marc Wilkins from Sydney as the "**prote**in complement of the gen**ome**", resulting in the birth of the field of "proteomics". This rather simple

G. Agnetti, PhD, FAHA (✉)
Laboratory of Cardiovascular Biochemistry, Division of Cardiology,
Department of Medicine, The Johns Hopkins University School of Medicine,
Baltimore, MD, USA

Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna,
Bologna, Italy
e-mail: gagnett1@jhmi.edu

M.J. Dunn, BSc, PhD
Conway Institute, University College Dublin, Dublin, Ireland

moment defined a time in history when the emphasis would slowly but steadily move from the role of genes to that of proteins and their modifications as defining traits of the phenotype, both in health and disease. This "Cultural Revolution" was largely made possible by the remarkable technological advances in the field of protein biochemistry that were achieved in the latter part of the twentieth century. In 1975 Pat O'Farrell and other groups optimized the way to separate and detect over 1100 proteins in a single experiment using two-dimensional polyacrylamide gel electrophoresis (2DE) [1]. Combined with the rapid developments in protein and peptide mass spectrometry (MS), and the creation of protein databases which followed, these combined techniques enabled by the late 1990s the separation, quantification and identification of hundreds to thousands of proteins starting from a single biological sample. Indeed, the study co-Authored by Marc Wilkins the year following the 1994 Siena meeting inaugurated a new era in protein biochemistry [2].

As we write, we have recently celebrated the 21st birthday of the proteome and in retrospect, many things have dramatically improved. To mark this occasion, a Special Issue celebrating the 20th birthday of proteomics, dedicated to the memory of Vitaliano Pallini (the former Supervisor of the Functional Proteomics group at the Molecular Biology Department at the University of Siena, Italy), was recently published in the Journal of Proteomics [3]. We direct the interested reader to this special issue for historical references as several scientists that participated in the birth of proteomics contributed with their personal views and memories.

The "new field" of proteomics (i.e. the complex of technology used to investigate the proteome), which could be also termed *advanced protein biochemistry*, is the result of the incredible technological advances that have allowed protein biochemists to study the subjects of their investigation in a more efficient way. However, along with the technicalities, as often happens in scientific history, new concepts and ideas arose thanks to the availability of these new technologies. Much like the invention of the microscope allowed van Leeuwenhoek to see *animalcules*, which led to the discovery of pathogens, proteomic technologies have revealed an almost unimaginable complexity in terms of what are increasingly referred to as *proteoforms* [4]. Therefore, the impact of proteomics is not only technological, but also highly biological in nature. It is likely that the interest in genes and transcripts that captured the attention of the scientific world (and US Congress) in the 1980s, was also fueled by the relative facility by which nucleic acids can be amplified, a luxury that is still missing in the field of proteomics. However, thanks to the steady advances in protein separation technology, and more recently the implementation of novel MS configurations and protocols, the possibility of mapping the proteome, decorated with its countless post-translational modifications (PTMs), now seems within reach. The outstanding protein scientists who endured long periods of anonymity in the genomic era should not be forgotten as it is thanks to them that we are now capable of achieving all the wonderful things that proteomics is making possible. Thanks to the passion, hard work and dedication of these researchers, the variety and complexity of the proteome has finally emerged. This realization is having and will continue to exert a tremendous impact on modern life sciences and biomedicine and it should be born in mind that, as much as proteomics, more so than other global approaches (or –omics), relies heavily on technologies, the driving force which will propel proteomics in the future will be its impact on biological sciences and biomedicine.

In this chapter, we will highlight the scientific landmarks that made the study of the proteome possible, with specific reference to the cardiovascular system. It is our belief that a thorough understanding of where the proteomic world comes from will highlight the path to a prosperous future.

A Brief History of Proteomics

As mentioned in the introductory paragraph, the "proteomic revolution" could not have been possible without the technique, now 41 years old, which was carefully optimized by Pat O'Farrell in 1975 [1]. Although other groups implemented alternative versions of orthogonal gel separation, the 2DE protocol optimized by O'Farrell was arguably superior at the time it was published [5]. Indeed, by combining the best approaches available at the time and with an eye to the future, he was able to separate and visualize over a thousand proteins from *E. Coli*, while other groups could "only" see a few hundred proteins. Part of this success was made possible by the choice of isotopic labeling to boost sensitivity by about three orders of magnitude with respect to other groups who used protein staining (such as Coomassie Brilliant Blue) [5]. This technique, which is still in use today, is based on the combination of two independent electrophoretic separation methods applied to the sample in an orthogonal fashion. The first separation, commonly referred to as the *first dimension*, facilitates the separation of proteins according to their charge properties (isoelectric point, pI) by isoelectric focusing (IEF) under denaturing conditions. This is followed by the *second dimension*, which exploits the denaturing properties of charged detergents (typically sodium dodecyl sulphate, SDS) in order to separate proteins based on their molecular mass (M_r) (Fig. 1.1, see Chaps. 2 and 7 for details). Both electrophoretic steps are carried out using polyacrylamide gels (PAGE) with different properties that were optimized in the course of the following decade. In fact, another landmark discovery that allowed 2DE to become widely used was the advent of immobilized pH gradients (IPG). The idea, mastered by Pier
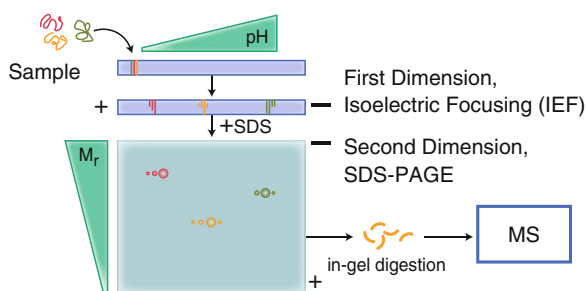

Fig. 1.1 Schematic of 2DE-MS. Proteins are first separated by 2DE according to their pI and molecular mass (M_r). The first separation in 2DE (IEF) is commonly referred to as "first dimension" whereas the second step (SDS-PAGE) is commonly referred to as "second dimension". After in-gel digestion proteins can be identified using MS (see text for details)

Giorgio Righetti and colleagues at the University of Milan (*Italy*) and developed within a large consortium involving several Universities and established scientists as well as an industrial partner (LKB at that time), led to the historical publication by Bjellqvist and colleagues in 1982 [6]. A few years later, based largely on optimization of the IPG 2DE technique in the laboratory of Angelika Gorg, the industrial partner (now Pharmacia) made the technology available at an industrial scale, enabling increased intra- and inter-lab reproducibility of 2DE [7] and the widespread diffusion of 2DE.

Using 2DE technology, Valerie Wasinger and colleagues were able to generate the data disclosed at the'94 Siena meeting and published the following year in the journal Electrophoresis [2]. In what is commonly referred to as the "first proteomic paper", these new technologies were combined with matrix-assisted laser desorption ionization–time of flight MS (MALDI-TOF MS, see Chap. 2), for the rapid identification of proteins using peptide mass fingerprinting (PMF) [8]. This MS-based approach is based on the digestion of the separated proteins with a protease (e.g. trypsin) and was utilized extensively in the early days of proteomics for the rapid identification of proteins. With the more recent popularity of tandem MS (see Chap. 2), allowing accurate determination of the sequence of peptides of the separated proteins, PMF was almost abandoned, only to make a glorious comeback as a rapid way to identify pathogens in biomedical labs in recent years [9] (see Chap. 18). The last key ingredient that was needed to identify proteins efficiently was the development of databases and algorithms to match the obtained MS spectra with protein names. Among the many contributions made by Norman and Leigh Anderson (father and son), the creation of one of the first successful algorithms to analyze 2DE maps [10] and the first online protein database (the Human Protein Index or HPI) [11] are two advances that accelerated the much needed creation of bioinformatic tools to mine the proteome (see Chaps. 12 and 14).

Operator Independence Days in Separation Techniques

Among the limitations of 2DE, the labor intensive and "artisanal" way they are performed paved the way to the success of more automatized techniques. To anybody who has successfully run a 2D gel, it is obvious how much training and effort is required to optimize both the conditions and the manual skills needed to obtain a well-resolved protein profile. With regard to the visual analysis of 2D gels and considering all of the effort involved in generating them, a parallel with visual arts is perhaps not too much a stretch of the imagination. The love-hate relationship of proteomic scientists with 2DE has made it something of a romantic journey, as testified for instance by the severe problems with the separation of membrane proteins, voiced so passionately by Thierry Rabilloud [12, 13].

Although several protein separation techniques have been implemented over the years, liquid chromatography (LC) became extremely popular for several reasons. Firstly its direct coupling with an MS instrument is relatively straightforward, decreasing the chance of contamination and avoiding a "transfer" step. In the origi-
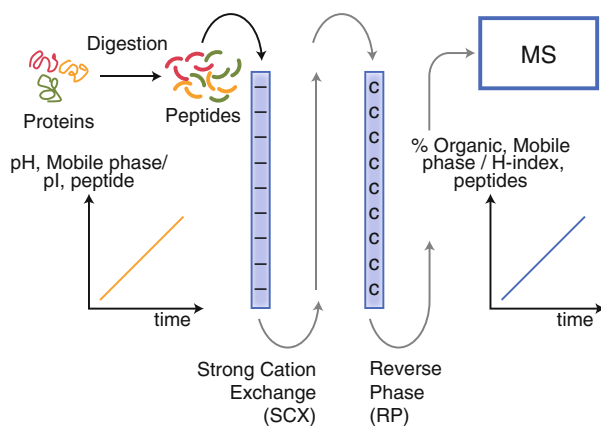
Fig. 1.2 Schematic of LC-MS. In "shotgun" proteomics proteins are first digested into peptides, which are generally more stable and soluble. Peptides can be then separated by liquid chromatographic techniques. Classically, Strong Cation Exchange (SCX) and Reverse Phase (RP) chromatography were combined in series to improve resolution. Liquid chromatography can be physically coupled to MS to identify the resulting peptides (see text and Chap. 4 & 7 for details)

nal iteration of what is referred to today as the "shotgun" approach, John Yates and colleagues also applied an "orthogonal" peptide separation approach by combining strong cation exchange (SCX) and reverse phase (RP) chromatography in sequence [14] (Fig. 1.2, see Chap. 4 & 7). This was in the late'90 and the limitations of shotgun approaches, including limited quantitation capabilities, also became evident in the following years. The use of isotopic metabolic or post- labeling, was successfully applied to improve the quantitation capabilities of LC-MS approaches in the decade that followed [15]. Briefly, labeling proteins with isotopic aminoacids (dividing cells [16]) or tags (post-mitotic primary cultures, tissues, etc. [17]), allowed resolution of peaks originating from the same peptide but deriving from different samples in the MS as a mass shift, in a highly quantitative fashion. Several iterations of these reagents resulted in a technique that is now both robust and commercially available (see Chap. 11) [18].

In a similar fashion, at the end the'90s a labeling approach that improved sensitivity, dynamic range and throughput in 2DE was developed by scientists at Carnegie Mellon in Pittsburgh [19]. This technology, currently commercialized by GE Healthcare, involves the pre-electrophoretic fluorescent labeling of protein samples with *N*-hydroxy-succinimidyl-ester derivatives of fluorescent cyanine (Cy) dyes and is known as two-dimensional difference gel electrophoresis (DIGE). This approach has the advantage that a pair of protein samples can be labeled separately with differently fluorescing Cy3 and Cy5 derivatives. The two samples can be mixed and then separated together on the same 2D gel. The resulting 2D gel is then scanned to acquire the Cy3 and Cy5 images separately using a fluorescent laser scanner. Furthermore, a sample labeled with a third Cy2 dye can be run on each 2D gel and used to normalize the signal among different gels. This DIGE approach dramatically reduces technical variability and exploits the high dynamic range of fluorescent staining for accurate quantitation. The issue of mul-

tiple proteins within the same 2DE spot still remains and, therefore, downstream validation of the changes observed in protein levels by other techniques is still essential. In addition, the relatively high price of the dyes and detection systems (a laser scanner able to detect the fluorescent probes) remain limiting factors.

As a result of the enormous advances in MS technology, it is nowadays possible to create an *in silico*, "unbiased" map of the proteome with little sample manipulation prior to MS [20, 21]. The combination of data independent acquisition (DIA) and targeted methods, such as multiple reaction monitoring (MRM [22]), now facilitate the accurate quantitation of peptides in the absence of isotopic labeling and in a proteome-wide fashion (see Chap. 10 on DIA/SWATH). The higher throughput capabilities and automatization of LC *versus* 2DE, combined with the increased capacity of detecting membrane and basic proteins further contributed to the popularity of LC-MS in the following decade and through to the present day. However, one limitation of shotgun approaches that remains to this day is that proteins are first digested into peptides, which are more stable, prior to separation by LC. From the standpoint of a biologist it is easy to understand how, by operating at the peptide level from the outset, important information can be lost. Indeed, so-called "top-down" approaches, which allow the analysis of whole proteins by MS (as opposed to "bottom-down" approaches, such as shotgun proteomics), are becoming increasingly popular (see Chap. 8 and [23]). This is also a consequence of the enormous amount of data and information generated by a single shotgun experiment, which can be challenging to store and analyze. The divide between bottom-up and top-down approaches is a blurred one and in fact some argue that 1D (SDS-PAGE or IEF) and 2DE techniques may be right in the middle, offering the opportunity to separate, or at least reduce the complexity of protein samples to a few intact proteins per spot. The capacity to see gross changes in molecular weight (e.g. degradation or proteolysis), or isoelectric point (e.g. phosphorylation) prior to digestion, combined with the natural propensity of the human brain to recognize patterns, suggest that 2DE may still have a role to play as a balance between top-down and bottom-up approaches, while celebrating its forty-first birthday [24].

Over a Century of Mass Spectrometry

Despite MS having a dominant role in proteomics these days, it is not a new technique. Indeed the first reports of the use of a very distant cousin of a modern MS date back to the work of Joseph John Thomson sometime around the turn of the twentieth century. Unbeknown to this young theoretical physicist at Cambridge University, MS would turn into, arguably, the most revolutionary technology in biomedicine over a century later. This embryonic MS, then called a parabola spectrograph, was initially used to investigate the very intimate components of matter, for which Thomson received a Nobel Prize in 1906 for "discovering the electron" [25].

It would take seven decades for MS to be able to be used to sequence proteins thanks to the work and dedication of Klaus Biemann and colleagues, right around

the time when protein biochemists were optimizing 2DE [26]. Indeed, the possibility of analyzing large organic molecules using MS was limited by the ability to convert them into ions that would be able to "fly" in the instrument, a requirement for MS analysis. MS was initially applied to smaller molecules such as metabolites or pharmaceuticals, which could be ionized without being fragmented during the process. The way that an MS instrument works can be equated to that of a very precise molecular scale, as we will see in details in the following Chapter. To ionize large organic molecules represented a major challenge in MS history, because the kinetic energy that needs to be transferred to the peptides for them to enter into a gas phase as ions and be separated using magnetic selectors, would cause their fragmentation.

It took the work of several groups and two Nobel prizes to overcome this limitation in the late '80s. Two different approaches were pursued starting from either solid (crystallized) or liquid samples. Both proved to be effective in the end and generated the two sources that are still in use today: electrospray ionization (ESI), and matrix-assisted laser desorption ionization (MALDI). As we will learn in the next Chapter an MS instrument can be divided in three sections: a source, which converts peptides into ions in a gas phase; an analyzer (or multiple analyzers in series, as for tandem MS); and a detector to "count" the peptides ions and/or their fragments, separated according to their mass (mass over charge or m/z to be precise) by the analyzer/s (Fig. 1.3).

The Nobel prizes were ultimately awarded to John Fenn (ESI) and Koichi Tanaka (MALDI) in 2002, for "their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecuels [27]". A little known fact about MALDI, is that although the Nobel prize was awarded to Dr. Tanaka and his group for their pioneering work on ionizing organic macromolecules from solid phase, the MALDI source, as it is currently used, was developed by two German scientists, Michael Karas and Franz Hillenkamp, who also named the technique [25]. This is one of the first examples of MS applied to the cardiovascular realm as the two scientists had a specific interest in mapping Ca^{2+} stores in
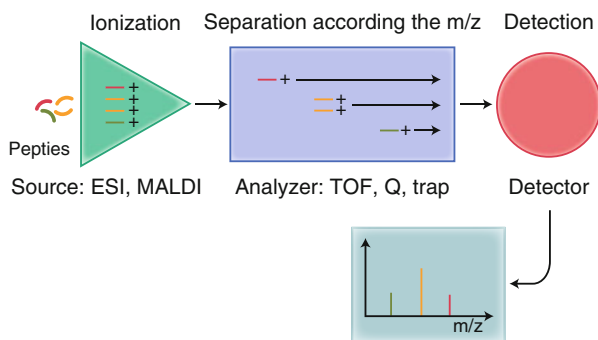

Fig. 1.3 Anatomy of a mass spectrometer. A mass spectrometer is best described by its "anatomical" components: a source, one or more analyzers and a detector (see text and Chap. 2 for details)

cardiac cells [25]. These "soft" ionization methods were first applied to nucleic acids. As we will see in the next paragraph, along with the creation of genomic database, nucleic acid research indirectly contributed to the development of proteomics. Lastly, as we will see in Chapter 2, several advances in MS, including the creation of new analyzers (such as orbital traps [28]) and improved fragmentation approaches [29], have remarkably enhanced the capability of modern instruments to the point that several thousands of proteins and their modifications can be accurately resolved nowadays.

Genomic and Proteomic Databases: From Genes to Proteins, and Back

As mentioned, one of the technologies that was developed during the genomic era and which served the proteomic cause very well was the creation of protein databases. In the first proteomic studies, proteins were identified using chemical Edman degradation and/or genetic tools which allowed scientists to assign a peptide to a protein, by translating it into a genetic sequence and matching it to those available in a particular genome. Since the very beginning of protein sequencing by MS, this new approach helped to identify open reading frames (ORFs) or sequenced genes thus perfecting their publication and annotation [30].

As mentioned above, the creation of protein databases to mine protein sequences was the result of the renaissance mind of the Andersons who created the first online protein database (HPI) [11] and the many implementations which followed. The sequencing of the genome of many organisms, including the human genome, allowed proteomic scientists to build protein databases predicted on the basis of the genetic info and ORFs. On this issue it is interesting to see that proteomic studies helped to find limitations in genomic databases [31]. The quest to match peptide spectra was greatly facilitated by the creation of *search and retrieval systems* (SRSs), algorithms, such as MASCOT [32], capable of scanning through a large database and returning the likelihood of a true-positive match between the experimental mass spectrum and a protein sequence listed *in silico*. As can be seen the sequencing of genomes not only allowed proteomics to flourish but for the former to "pay back" by pinpointing limitations in the algorithms used to predict and annotate genes in the respective databases. Indeed, despite genetic information being there it is important to confirm that it is relevant to the phenotype, or expressed. Even when that is the case, the predictive value of genes is highly limited by mRNA stability, alternative splicing and post-translational modifications [33]. Large-scale proteomic studies are inconceivable without the aid of bioinformatics and this finally allowed us to break free from the "one gene to one protein" dogma [34]. It is our hope that, as our technologies and bioinformatic tools perform more accurately and errors are corrected, the detailed picture of the molecular phenotype (aka the proteome) will finally reach a sublime resolution.

Proteomics with a Heart

We already mentioned how, thanks to the work of Michael Karas and Franz Hillenkamp in Frankfurt, cardiovascular research was involved with cutting-edge proteomic research early on in the history of proteomics. There are many unanswered questions in the cardiovascular realm many of which have a deep impact on public health due to the widespread incidence of a multiplicity of cardiovascular diseases [18]. Proteomics would be a natural tool to investigate mechanisms, generate new hypotheses and test the predictive value of novel biomarkers. We are lucky to have participated in the development of cardiovascular proteomics. Indeed, the first studies on the heart using 2DE anticipated the 1995 publication by Valerie Wasinger and colleagues of a much more unpleasant proteome [2]. Mike Dunn and Peter Jungblut were among the first scientists utilizing these methods to study the human heart, and published the first human heart 2D maps the same year the proteome was "born" [35, 36]. Shortly after, the same groups utilized MS to identify numerous myocardial proteins from 2DE [8, 37]. In their seminal report from 1998, Mike Dunn and colleagues compared cardiac biopsies from dilated cardiomyopathy (DCM) and ischemic heart disease (IHD) patients and found several significant changes that were also monitored at the isolated cell level for fibroblasts, mesothelial and endothelial cells, and cardiac myocytes [38]. To our knowledge this is the first report of a proteomic study addressing cardiovascular disease (Fig. 1.4). Interestingly, the changes in desmin and the chaperone alpha-B-crystallin first reported over 25 years ago are still the object of an intense investigation by several scientists (including our group [39]). Specifically, the idea that desmin can form preamyloid oligomers in the heart, similar to those found in Alzheimer's and Parkinson's diseases as well as several other proteinopathies, suggest that protein misfolding could play a major role in many diseases affecting the majority of the population in westernized societies, including cardiovascular disease. The role of
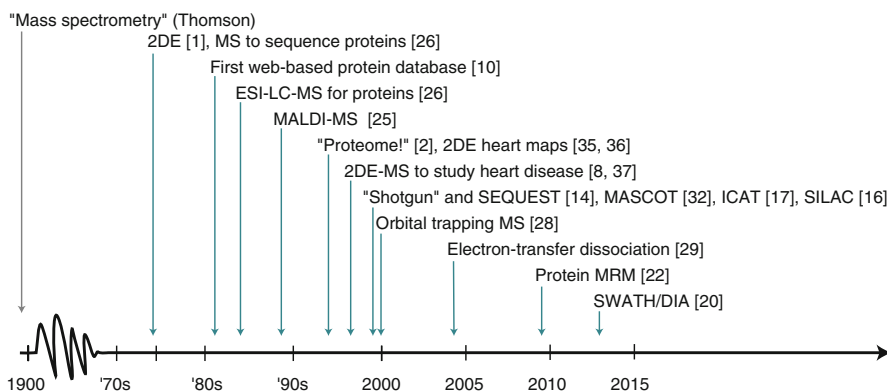

Fig. 1.4 A Timeline for the history of cardiovascular proteomics. See text for details
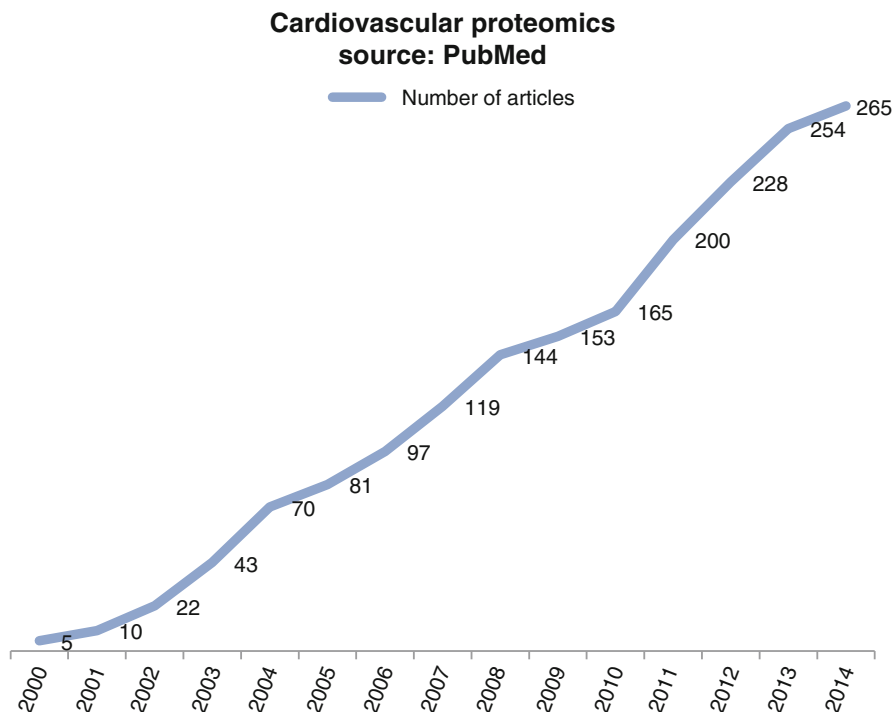
**Cardiovascular proteomics
source: PubMed**

Fig. 1.5 Cardiovascular proteomics articles in the last 15 years. Plot displaying the number of articles identified by the search term "cardiovascular proteomics" per year since 2000 (Source, PubMed)

posttranslational modifications in this process is also rapidly emerging [40]. Therefore, proteomics may also help to generate new views and cutting edge idea in cardiovascular research.

With the new century, proteomics started to feature rapidly in the cardiovascular field. More pioneers of the caliber of Jennifer Van Eyk and Peipei Ping, started exploiting these emerging technologies to address a number of different aspects of cardiovascular disease spanning from signaling [36]. to biomarker discovery [41]. Thanks to the pioneering work of these scientists and others the field grew at an almost exponential pace, scoring a total number of 2011 "cardiovascular proteomics" publications at the time of writing. Although this does not include all of the work done in "specialty areas" and is limited by the searched terms, it does provide a reliable estimate of the upward trend of the field for the last 15 years (Fig. 1.5). The term was used in only five articles in 2000, and in 265 articles in 2014. The appeal of using chromatography over 2DE gels has also rapidly increased and several comparative studies describing the complementary use of the different approaches to achieve more extensive coverage have been published in the last decade. As proteomics evolves, and that is happening very quickly, we are sure that this approach will be increasingly incorporated into studies addressing the main unsolved questions concerning prevention, diagnosis and cure of cardiovascular disease.

Historical Crossroads: Pick Your New Technologies Wisely

Once proteomics became available to cardiovascular scientists at the beginning of this century, an immediate challenge came with it: what technology should they use to address a specific biomedical question? This is true for both basic and translational studies. The number of technological platforms also exploded. To keep up with a continuously evolving field is not an easy task, especially when it is as technologically based as proteomics. In our opinion, one of the responsibilities of the proteomic community is to explain these aspects to the ever-increasing number of collaborators in a way that is unbiased and comprehensible. Untimely, there is a tendency in some scientific environments to claim potential benefits even if they are only on the horizon of a particular technology. The increasing pressure to obtain research funding at a time when the economy is "breathing slowly" may accentuate this tendency.

There are several examples of why it is best to under-promise and over-deliver than *vice versa*. Perhaps one of such example that should be born in mind when establishing a collaborative effort or investing in a new technology is the experience with SELDI (surface-enhanced laser desorption ionization). This is a modified and patented version of MALDI that became quite popular in the early days of proteomics. Despite the promise of accelerating the discovery of new biomarkers, when it was first released, the technology was based on a low resolution MALDI-TOF instrument with a proprietary technology to fractionate the samples directly onto the MALDI (or SELDI) plate, the support that allows samples to enter the instrument. Despite the initial "hype" that seems to accompany most new technologies, the use of SELDI has now rapidly declined, probably because the method did not meet initial expectations, possibly due to the high impact of matrix effects, limited reproducibility and limited identification capabilities [42].

This is just one of the many examples that are a feature of the history of proteomics, and probably of many other fields. Another lesson learned is that the SELDI instrument was originally packaged in a "black box" type of format, claiming a simple sample in/data out operation. This is a very important lesson for new and old proteomic scientists alike. The field is complicated, and as much as there is a purpose in trying to simplify technologies in order to allow their broader use, certain aspects simply cannot be simplified. When a technology does not deliver what it had promised, the impact on its future tends to be more dramatic, possibly because there is a collective memory in the scientific community, which tends to adjust to new concepts with a certain delay. It is also possible that this fact alone may limit the chances of "redemption" for technologies that are initially advertised to address more than can be expected from them. The "market" of proteomic technologies is a bit like the stock market in that it is extremely volatile, and at times expensive. For proteomic beginners it may be wise to rely on the advice of more experienced scientists in the field, and since we all have our passions and beliefs, perhaps diversifying the portfolio of "proteomic brokers" could also prove beneficial. Last but not least, proteomic scientists need to make themselves understandable to the general scientific community, avoiding technical jargon when possible and sharing their unbiased knowledge. This book represents an effort in this very direction.

What Do We Leave Behind as We Move Forward: Bringing the Fun Back into Science

This is a tremendously exciting moment in science, including the cardiovascular field. The capabilities offered by the latest generation of proteomic technologies are almost infinite. We can see most of the proteome, quantify it and finally make sense of it. Thinking back on how hard previous generations of scientists had to work in order to optimize these technologies should make us feel extremely lucky. It took decades to couple an LC with an MS, and decades to analyze peptides and proteins using it. It took decades to sequence the genomes that are available today, create the corresponding protein databases and the algorithms to exploit them to generate proteomic data. It almost appears that everything had to be optimized for the current generation of proteomic scientists to "have fun with it". Proteomics can be painful, as science at large has the tendency to be sometimes. However, it has the potential and capability to bring very diverse expertise around the same table: statisticians, engineers, biochemists and physicians can find some exciting, common ground to move science and medicine forward (hopefully, as is the case with the present book). With this also comes the responsibility to actually change medicine and translate the impact of all these rich technological gifts back into some good for human health. There is an increasing demand for technical expertise and scientists that can translate the difficult languages of math, engineering, protein biochemistry and medicine. It takes time and dedication to have enough knowledge of these languages to be able to contribute something to the table. However, it is not inconceivable that proteomics will become a curriculum in many undergraduate and graduate programs in its own right, as happened previously for biotechnology, molecular biology, etc.. In all, these opportunities for collective growth lie just ahead of us waiting to be picked like ripe fruit.

With such opportunities in prospect, perhaps the best way to move forward is to honor those that took us this far by pursuing the best science possible. Of course quality and integrity are two pillars of science, or at least should be. As we have learned in this chapter, time will tell if what we have fulfilled these two core principles. Moreover, in an era where communication and data access is so rapid it may become increasingly easier to get distracted and lose focus of the scientific goals. In addition, when being exposed to technologies, there is also a risk of falling in love with them. In summary, with all these novelties comes the risk of forgetting the knowledge generated by other means and during earlier times, when perhaps it took longer to separate proteins or to analyze them, but this time could be used to learn about the functional aspects of our discoveries, in-depth. Many proteomic scientists in the cardiovascular field had to write scientific articles explaining why a list of proteins, many of which we had not heard about before, change in a coordinated fashion. What has been published about the role of these mysterious proteins and the biological effect of their known post-translational modifications that we see changing in a particular disease state are very important aspects of our research. To us, another good way to honor the work of the fathers of proteomics would simply be to "do our homework", try to put our observations in the context of what is

already known without the need to rewrite science. Lastly, as we stand on the shoulders of the scientists who made us who we are, we honor their legacy by learning from their mistakes and endeavoring to match their successes.

References

1. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem. 1975;250:4007–21.
2. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: mycoplasma genitalium. Electrophoresis. 1995;16:1090–4.
3. Bini L, Calvete JJ, Turck N, Hochstrasser D, Sanchez J. Special Issue: "20 years of Proteomics" in memory of Vitaliano Pallini. J Proteomics. 2014;107:1–146.
4. Smith, L.M., Kelleher, N.L., and Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. Nat Methods. 2013;10:186–7.
5. Righetti PG. The Monkey King: a personal view of the long journey towards a proteomic Nirvana. J Proteomics. 2014;107:39–49.
6. Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W. Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods. 1982;6:317–39.
7. Gorg A, Drews O, Luck C, Weiland F, Weiss W. 2-DE with IPGs. Electrophoresis. 2009;30 Suppl 1:S122–32.
8. Sutton CW, Pemberton KS, Cottrell JS, Corbett JM, Wheeler CH, Dunn MJ, Pappin DJ. Identification of myocardial proteins from two-dimensional gels by peptide mass fingerprinting. Electrophoresis. 1995;16:308–16.
9. Randell P. It's a MALDI but it's a goodie: MALDI-TOF mass spectrometry for microbial identification. Thorax. 2014;69:776–8.
10. Anderson NL, Taylor J, Scandora AE, Coulter BP, Anderson NG. The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. Clin Chem. 1981;27:1807–20.
11. Anderson NG, Anderson L. The Human Protein Index. Clin Chem. 1982;28:739–48.
12. Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: un amour impossible? Electrophoresis. 2000;21:1054–70.
13. Rabilloud T. Membrane proteins and proteomics: love is possible, but so difficult. Electrophoresis. 2009;30 Suppl 1:S174–80.
14. Yates 3rd JR, Carmack E, Carmack E, Hays L, Link AJ, Eng JK. Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. Methods Mol Biol. 1999;112:553–69.
15. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nat Genet. 2003;33(Suppl):311–23.
16. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A. 1999;96:6591–6.
17. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17:994–9.
18. Agnetti G, Husberg C, Van Eyk JE. Divide and conquer: the application of organelle proteomics to heart failure. Circ Res. 2011;108:512–26.
19. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis. 1997;18:2071–7.