# Proteomics

Guest Editors: Benjamin A. Garcia, Helen J. Cooper,
Pieter C. Dorrestein, Kai Tang, and Beatrix M. Ueberheide

Proteomics

Proteomics

Guest Editors: Benjamin A. Garcia, Helen J. Cooper,
Pieter C. Dorrestein, Kai Tang, and Beatrix M. Ueberheide

Biophysics

Miguel Castanho, Portugal
P. Bryant Chase, USA
Kuo-Chen Chou, USA
Rizwan Khan, India

Ali A. Khraibi, Saudi Arabia
Rumiana Koynova, USA
Serdar Kuyucak, Australia
Jianjie Ma, USA

S. B. Petersen, Denmark
Peter Schuck, USA
Claudio M. Soares, Portugal

Cell Biology

Ricardo Benavente, Germany
Omar Benzakour, France
Sanford I. Bernstein, USA
Phillip I. Bird, Australia
Eric Bouhassira, USA
Mohamed Boutjdir, USA
Chung-Liang Chien, Taiwan
Richard Gomer, USA
Paul J. Higgins, USA
Pavel Hozak, Czech Republic

Xudong Huang, USA
Anton M. Jetten, USA
Seamus J. Martin, Ireland
Manuela Martins-Green, USA
Shoichiro Ono, USA
George Perry, USA
Mauro Piacentini, Italy
George E. Plopper, USA
Lawrence Rothblum, USA
Ulrich Scheer, Germany

Michael Sheetz, USA
James L. Sherley, USA
G. S. Stein, USA
Richard Tucker, USA
Thomas van Groen, USA
Andre Van Wijnen, USA
Steve Winder, UK
Chuanyue Wu, USA
Bin-Xian Zhang, USA

Genetics

Adewale Adeyinka, USA
Claude Bagnis, France
James Birchler, USA
Susan Blanton, USA
Barry J. Byrne, USA
Ranajit Chakraborty, USA
Domenico Coviello, Italy
Sarah H. Elsea, USA
Celina Janion, Poland
J. Spencer Johnston, USA

M. Ilyas Kamboh, USA
Feige Kaplan, Canada
Manfred Kayser, The Netherlands
Brynn Levy, USA
Xiao Jiang Li, USA
Thomas Liehr, Germany
James M. Mason, USA
Mohammed Rachidi, France
Raj S. Ramesar, South Africa
Elliot D. Rosen, USA

Dharambir K. Sanghera, USA
Michael Schmid, Germany
Markus Schuelke, Germany
Wolfgang A. Schulz, Germany
Jorge Sequeiros, Portugal
Mouldy Sioud, Norway
Meena Upadhyaya, UK
Rongjia Zhou, China

Genomics

Vladimir Bajic, Saudi Arabia
Margit Burmeister, USA
Settara Chandrasekharappa, USA
Yataro Daigo, Japan
J. Spencer Johnston, USA

Vladimir Larionov, USA
Thomas Lufkin, Singapore
Joakim Lundeberg, Sweden
John L. McGregor, France
John V. Moran, USA

Yasushi Okazaki, Japan
Gopi K. Podila, USA
Paul B. Samollow, USA
Momiao Xiong, USA

Immunology

Hassan Alizadeh, USA
Peter Bretscher, Canada
Robert E. Cone, USA
Terry L. Delovitch, Canada
Anthony L. DeVico, USA
Nick Di Girolamo, Australia
Dorothy Mark Estes, USA
Soldano Ferrone, USA
Jeffrey A. Frelinger, USA
John Gordon, UK
John Robert Gordon, Canada

James D. Gorham, USA
Silvia Gregori, Italy
Thomas Griffith, USA
Young S. Hahn, USA
Dorothy E. Lewis, USA
Bradley W. McIntyre, USA
R. Lee Mosley, USA
Marija Mostarica-Stojković, Serbia
Hans Konrad Muller, Australia
Ali Ouaissi, France
Kanury V. S. Rao, India

Yair Reisner, Israel
Harry W. Schroeder, USA
Wilhelm Schwaeble, UK
Nilabh Shastri, USA
Yufang Shi, China
Piet Stinissen, Belgium
Hannes Stockinger, Austria
J. W. Tervaert, The Netherlands
Graham R. Wallace, UK

Microbial Biotechnology

Jozef Anné, Belgium
Yoav Bashan, Mexico
Marco Bazzicalupo, Italy
Nico Boon, Belgium
Luca Simone Cocolin, Italy

Peter Coloe, Australia
Daniele Daffonchio, Italy
Han de Winde, The Netherlands
Yanhe Ma, China
Bernd Rehm, New Zealand

Angela Sessitsch, Austria
Effie Tsakalidou, Greece
Juergen Wiegel, USA

Microbiology

David Beighton, UK
Steven R. Blanke, USA
Stanley Brul, The Netherlands
Isaac K. O. Cann, USA
Peter Dimroth, Switzerland
Stephen K. Farrand, USA
Alain Filloux, UK

Gad Frankel, UK
Roy Gross, Germany
Hans-Peter Klenk, Germany
Tanya Parish, UK
Gopi K. Podila, USA
Frederick D. Quinn, USA
Didier Raoult, France

Isabel Sá-Correia, Portugal
Pamela L. C. Small, USA
Lori Snyder, UK
Michael Thomm, Germany
Henny van der Mei, The Netherlands
Schwan William, USA

Molecular Biology

Rudi Beyaert, Belgium
Michael Bustin, USA
Douglas Cyr, USA
Kostas Iatrou, Greece
Lokesh Joshi, Ireland
David W. Litchfield, Canada

Noel F. Lowndes, Ireland
Wuyuan Lu, USA
Patrick Matthias, Switzerland
John L McGregor, France
Sherry Mowbray, Sweden
Elena Orlova, UK

Yeon-Kyun Shin, USA
William S. Trimble, Canada
Lisa Wiesmuller, Germany
Masamitsu Yamaguchi, Japan

Virology

Contents

Editorial

Proteomics

**Benjamin A. Garcia,[1] Helen J. Cooper,[2] Pieter C. Dorrestein,[3]
Kai Tang,[4] and Beatrix M. Ueberheide[5]**

[1] *Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA*
[2] *School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[3] *Departments of Pharmacology, Chemistry and Biochemistry, Skaggs School of Pharmacy and Pharmaceutical Sciences,
   University of California, San Diego, La Jolla, CA 92093, USA*
[4] *School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, Singapore 63755*
[5] *The Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA*

Correspondence should be addressed to Benjamin A. Garcia, bagarcia@princeton.edu

Mass spectrometry-based proteomics has played an increasing role in biological and biomedical research as gauged by the ever increasing numbers of thousands of publications arising each year. With advances in proteomic technology, bioinformatics and different experimental approaches, sampling of near complete proteomes for some organisms, assessment of entire signaling pathways, and discovery of biomarkers of disease have become possible. The broad range and ease of implementation of various workflows combined with assess to robust instrumentation has allowed for continued crossover of proteomics to many diverse biological and even clinical areas. Here, the Journal of Biomedicine and Biotechnology presents the "Proteomics 2011" special issue. Although only representing a small fraction of the contribution of mass spectrometry to many research fields, the articles presented here definitely encompass the excitement that proteomics research can provide to complement traditional biological studies.

To start this special issue, we have five contributions that review several aspects of proteomic research. W. Zhu et al. discuss the different contributing factors for accurate label-free quantification, while challenges that remain for proteomic SELDI experiments to be successful in biomarker discovery are covered by M.-P. Merville and coworkers. Two more specialized reviews cover parts of the biology of translation in eukaryotes mediated by internal ribosome entry sites and intrinsically disordered protein structures (Pecheco et al. and V. N. Uversky, resp.). Lastly, how

proteomics may play a role in crop disease prevention by understanding the pathogenesis of fungi is covered by J. V. Jorrín-Novo and coworkers.

Articles describing the development of methods and computational approaches are also well represented in this special issue. On the methods side, we have J. Hou et al. demonstrating improved MALDI-MS of phosphopeptides by an optimized matrix. Improved sample preparations including separations are utilized by M. Centlow et al. and E. Gamelin and coworkers to uncover potential biomarkers of preeclamptic placenta and to improve proteome coverage of plasma, respectively. 2D liquid chromatography coupled to mass spectrometry is used to determine proteomic response to cadmium stress in plants by G. Visioli et al. Identifying released bioactive proteins from human platelets is addressed by G. Cagney and coworkers. Dealing with the computational end of proteomic studies, T. H. Corzett et al. describe statistical measures for determining variation resulting from quantification of the human plasma proteome separated by 2D difference gel electrophoresis. Enhanced peptide and protein identification is addressed by L. Pattini and coworkers who present a novel method for preprocessing proteomic data, and by K. Macur et al. who utilize peptide retention time as an extra indication of false positives.

Lastly, we showcase the application of mass spectrometry-based proteomics for the determination of potential biomarkers from various biological and clinical conditions. J. Svasti and coworkers investigate the secretomes from

different carcinoma cell lines by GeLC-MS to find potential tumor biomarkers. In search of an elusive Downs syndrome (DS) protein marker, S. Hahn and coworkers use iTRAQ labeling to screen plasma from DS pregnancies. Campostrini et al. look to quantify the iron regulating hepcidin protein isoforms in patients afflicted by chronic haemodialysis. Overall, we hope this special issue brings to light the potential that proteomic experiments possess and the impacts that could be made to health and disease.

<div align="right">

Benjamin A. Garcia
Helen J. Cooper
Pieter C. Dorrestein
Kai Tang
Beatrix M. Ueberheide

</div>

Review Article

Mass Spectrometry-Based Label-Free Quantitative Proteomics

Wenhong Zhu,[1] Jeffrey W. Smith,[1] and Chun-Ming Huang[2]

[1] *Center on Proteolytic Pathways, Burnham Institute for Medical Research, 10901 N. Torrey Pines Road, La Jolla, CA 92037, USA*
[2] *Department of Dermatology, University of California, San Diego, CA 92093, USA*

Correspondence should be addressed to Wenhong Zhu, wzhu@burnham.org and Chun-Ming Huang, chh001@ucsd.edu

In order to study the differential protein expression in complex biological samples, strategies for rapid, highly reproducible and accurate quantification are necessary. Isotope labeling and fluorescent labeling techniques have been widely used in quantitative proteomics research. However, researchers are increasingly turning to label-free shotgun proteomics techniques for faster, cleaner, and simpler results. Mass spectrometry-based label-free quantitative proteomics falls into two general categories. In the first are the measurements of changes in chromatographic ion intensity such as peptide peak areas or peak heights. The second is based on the spectral counting of identified proteins. In this paper, we will discuss the technologies of these label-free quantitative methods, statistics, available computational software, and their applications in complex proteomics studies.

1. Introduction

Mass spectrometry plays a central role in proteomics [1]. In addition to global profiling of the proteins present within a system at a given time, information on the level of protein expression is increasingly required in proteomics studies [1, 2]. Protein separation and comparison by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), followed by mass spectrometry (MS) or tandem mass spectrometry (MS/MS) identification is the classical method for quantitative analysis of protein mixtures [3]. In this method, the intensity of the protein stain is used to make a determination regarding the quantity of a particular protein. The development of 2D Fluorescence Difference Gel Electrophoresis (2D-DIGE) gives more accurate and reliable quantification information of protein abundance because the samples to be compared are run together on the same gel, eliminating potential gel-to-gel variation [4]. However, spots on a given 2D gel often contain more than one protein, making quantification ambiguous since it is not immediately apparent which protein in the spot has changed. In addition, any 2D gel approach is subject to the restrictions imposed by the gel method, which include limited dynamic range, difficulty handling hydrophobic proteins, and difficulty detecting proteins with extreme molecular weights and pI values.

The development of non-gel-based, "shotgun" proteomic techniques such as Multidimensional Protein Identification (MudPIT) has provided powerful tools for studying large-scale protein expression and characterization in complex biological systems [5, 6]. Non-gel-based quantitative proteomics methods have, therefore, also been developed significantly in recent years. Because the chemical and physical properties of isotope labeled compounds are identical to properties of their natural counterparts except in mass, isotope labeled molecules were incorporated into mass spectrometry-based proteomics methods as internal standards or relative references. A number of stable isotope labeling approaches have been developed for "shotgun" quantitative proteomic analysis. These include Isotope-Coded Affinity Tag (ICAT), Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), ^{15}N/^{14}N metabolic labeling, ^{18}O/^{16}O enzymatic labeling, Isotope Coded Protein Labeling (ICPL), Tandem Mass Tags (TMT), Isobaric Tags for Relative and Absolute Quantification (iTRAQ), and other chemical labeling [7, 8]. These stable isotope labeling methods have provided valuable flexibility while using quantitative proteomic techniques to study protein changes in complex samples. However, most labeling-based quantification approaches have potential limitations. These include increased time and complexity of sample preparation, requirement for higher sample concentration, high cost of the reagents, incomplete labeling,
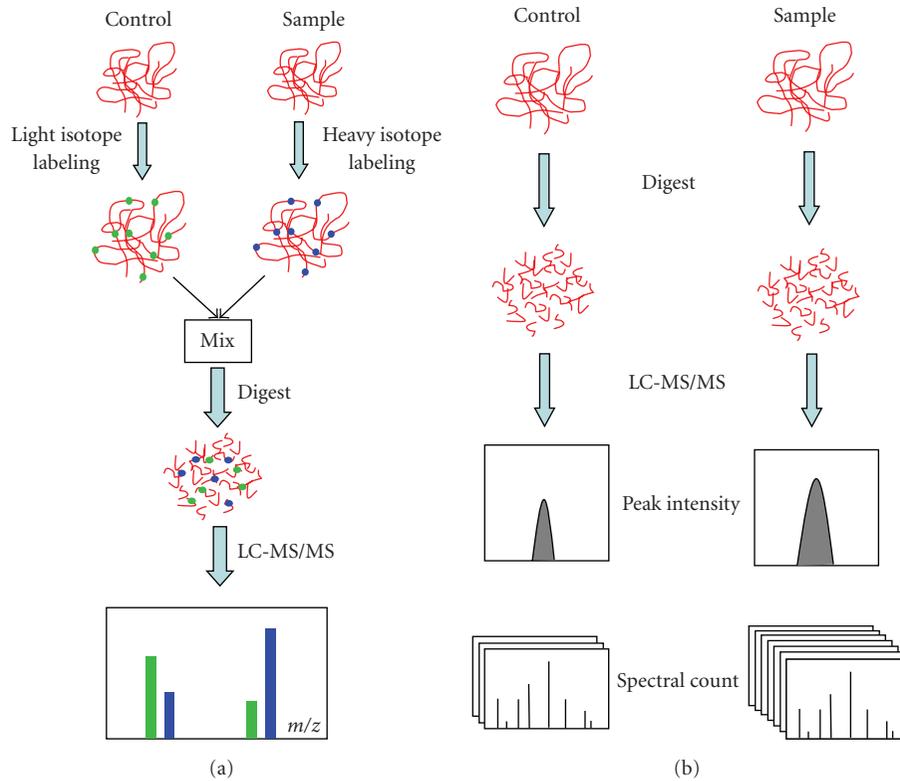
FIGURE 1: General approaches of quantitative proteomics. (a) Shotgun isotope labeling method. After labeling by light and heavy stable isotope, the control and sample are combined and analyzed by LC-MS/MS. The quantification is calculated based on the intensity ratio of isotope-labeled peptide pairs. (b) Label-free quantitative proteomics. Control and sample are subject to individual LC-MS/MS analysis. Quantification is based on the comparison of peak intensity of the same peptide or the spectral count of the same protein.

and the requirement for specific quantification software. Moreover, so far only TMT and iTRAQ allow the comparison of multiple (up to 8) samples at the same time. The other labeling methods can only compare the relative quantity of a protein between 2 and 3 different samples. There has, therefore, been increased interest in label-free shotgun proteomics techniques in order to address some of the issues of labeling methods and achieve faster, cleaner, and simpler quantification results [7, 9–11].

2. Label-Free Quantitative Proteomics

Regardless of which label-free quantitative proteomics method is used, they all include the following fundamental steps: (i) sample preparation including protein extraction, reduction, alkylation, and digestion; (ii) sample separation by liquid chromatography (LC or LC/LC) and analysis by MS/MS; (iii) data analysis including peptide/protein identification, quantification, and statistical analysis. In protein-labeling approaches, different protein samples are combined together once labeling is finished and the pooled mixtures are then taken through the sample preparation step before being analyzed by a single LC-MS/MS or LC/LC-MS/MS experiment (Figure 1(a)). In contrast, with label-free quantitative methods, each sample is separately prepared, then subjected to individual LC-MS/MS or LC/LC-MS/MS

runs (Figure 1(b)). Protein quantification is generally based on two categories of measurements. In the first are the measurements of ion intensity changes such as peptide peak areas or peak heights in chromatography. The second is based on the spectral counting of identified proteins after MS/MS analysis. Peptide peak intensity or spectral count is measured for individual LC-MS/MS or LC/LC-MS/MS runs and changes in protein abundance are calculated via a direct comparison between different analyses.

2.1. Relative Quantification by Peak Intensity of LC-MS. In LC-MS, an ion with a particular m/z is detected and recorded with a particular intensity, at a particular time. It has been observed that signal intensity from electrospray ionization (ESI) correlates with ion concentration [12]. The label-free quantification of peptide/protein via peak intensity in LC-MS was first studied by loading 10 fmol–100 pmol of myoglobin digests to nano-LC and analyzing by LC/MS/MS [13]. When the chromatographic peak areas of the identified peptides were extracted and calculated, the peak areas were found to increase with increased concentration of injected peptides. After the peak areas of all identified myoglobin peptides were combined and plotted against the protein amount, the peak area was found to correlate linearly to the concentration of protein (r^2 ℑ 0.991). The strong correlation between chromatographic peak areas and the

peptide/protein concentration remained when myoglobin was spiked into a complex mixture (human serum) and its digests were detected by LC-MS/MS. The results of quantitative profiling were further improved by normalizing the calculated peak areas [13, 14].

Although these early studies showed that the relative quantification of the peptides could be achieved via direct comparison of peak intensity of each peptide ion in multiple LC-MS datasets, applying this method for the analysis of changes in protein abundances in complex biological samples had some practical constraints. First, even the same sample can result in differences in the peak intensities of the peptides from run to run. These differences are caused by experimental variations such as differences in sample preparation and sample injection. Normalization is required to account for this kind of variation. Second, any experimental drifts in retention time and m/z will significantly complicate the direct, accurate comparison of multiple LC-MS datasets. Chromatographic shifts may occur as a result of multiple sample injections onto the same reverse-phase LC column. Unaligned peak comparison will result in large variability and inaccuracy in quantification. Thus, highly reproducible LC-MS and careful chromatographic peak alignment are required and critical in this comparative approach. Last, the large volume of data collected during LC-MS/MS analysis of complex protein mixtures requires the data analysis of these spectra to be automated. Therefore, capable computer algorithms were developed in the later studies in order to solve these issues and automatically compare the peak intensity data between LC-MS samples at a comprehensive scale. Several similar steps in data processing were performed in these label-free quantifications. Peptide peaks were first distinguished from background noise and from neighboring peaks (peak detection). Isotope patterns were assigned by deconvolution. LC-MS retention times were carefully adjusted in order to correctly match the corresponding mass peaks between multiple LC-MS runs (peak matching). Chromatographic peak intensity (either peak area or peak height) was calculated and normalized to enable a more accurate matching and quantitation. Finally, statistical analysis such as Students t-test was performed to determine the significance of changes between multiple samples [15–17].

Automatic comparison of peak intensity from multiple LC-MS datasets is well suited for clinical biomarker discovery, which normally requires high sample throughput. The following studies were all performed using this label-free quantitative approach. The comparison of control and radiated human colon cancer cells proved the reproducibility of this label-free approach [18]; the serum proteomic profiling of familial adenomatous polyposis patients revealed multiple novel celecoxib-modulated proteins [19]; proteins significantly associated with metastasis were identified by analyzing paraffin-embedded archival melanomas [20]; the analysis of 55 clinical serum samples from schizophrenia patients and healthy volunteers identified hundreds of differentially expressed serum proteins [21]; diagnostic markers and protein signatures were recognized from the serum of Gaucher patients [22] and the cerebrospinal fluid of schizophrenia patients [23].

2.2. Relative Quantification by Spectral Count. In the spectral counting approach, relative protein quantification is achieved by comparing the number of identified MS/MS spectra from the same protein in each of the multiple LC-MS/MS or LC/LC-MS/MS datasets. This is possible because an increase in protein abundance typically results in an increase in the number of its proteolytic peptides, and vice versa. This increased number of (tryptic) digests then usually results in an increase in protein sequence coverage, the number of identified unique peptides, and the number of identified total MS/MS spectra (spectral count) for each protein [24]. Liu et al. studied the correlation between relative protein abundance and sequence coverage, peptide number, and spectral count. It was demonstrated that among all the factors of identification, only spectral count showed strong linear correlation with relative protein abundance (r^2 ℑ 0.9997) with a dynamic range over 2 orders of magnitude [25]. Therefore, spectral count can be used as a simple but reliable index for relative protein quantification. An intriguing study evaluated relative quantification of protein complex by spectral counting-based method and $^{15}N/^{14}N$ isotope labeled, ion chromatographic method [26]. The crude membrane proteins extracted from *S. cerevisiae* grown in rich and minimal media were analyzed by MudPIT and quantified using both approaches. It was found that the two quantitative methods showed a strong correlation when the peptides with high signal-to-noise ratio in the extracted ion chromatogram were used in the comparison. Moreover, spectral counting-based quantification is proved more reproducible and has a larger dynamic range than the peptide ion chromatogram-based quantification [26].

In contrast to the chromatographic peak intensity approach, which requires delicate computer algorithms for automatic LC-MS peak alignment and comparison, no specific tools or algorithms have been developed specially for spectral counting due to its ease of implementation. However, normalization and statistical analysis of spectral counting datasets are necessary for accurate and reliable detection of protein changes in complex mixtures. A simple normalization method based on total spectral counts has been reported to account for the variation from run to run [27]. Since large proteins tend to contribute more peptide/spectra than small ones, a normalized spectral abundance factor (NSAF) was defined to account for the effect of protein length on spectral count [28, 29]. NSAF is calculated as the number of spectral counts (SpC) identifying a protein, divided by the protein's length (L), divided by the sum of SpC/L for all proteins in the experiment. NSAF allows the comparison of abundance of individual proteins in multiple independent samples and has been applied to quantify the expression changes in various complexes [29, 30].

Five different statistical tests have been compared by Zhang et al. to evaluate the significance of comparative quantification by spectral counts [31]. The Fisher's exact test,

TABLE 1: Commercially available software for label-free analysis.

Software	Producer	Quantification	Website
Decyder MS	GE Healthcare	Peak intensity	http://www5.gelifesciences.com/
SIEVE	Thermo Electron	Peak intensity	http://www.thermo.com/
Elucidator	Rosetta	Peak intensity, spectral count	http://www.rosettabio.com/
ProteinLynx	Waters	Peak intensity	http://www.waters.com/

goodness-of-fit test (G-test), AC test, Student's t-test, and Local-Pooled-Error (LPE) test were performed on spectral count data collected by MudPIT analysis of yeast digests. The Student's t-test was found to be the best when three or more replicates are available. The Fisher's exact test, G-test, and AC test can be used when the number of replications is limited (one or two), while G-test has the advantage due to its computational simplicity.

Relative quantification by spectral count has been widely applied in different biological complex, including analysis of urine sample from healthy donors and patients with acute inflammation [32], finding biomarkers in human saliva proteome in type-2 diabetes [33], comparison of protein expression in yeast and mammalian cells under different culture conditions [11, 26, 29], distinguishing lung cancer from normal [34], screening of phosphotyrosine-binding proteins in mammalian cells [35], and identifying differential plasma membrane proteins from terminally differentiated mouse cell lines [36].

2.3. Absolute Label-Free Quantification. In addition to relative quantification, label-free proteomics methods can also be used in the determination of absolute abundance of proteins. Protein abundance index (PAI), defined as the number of identified peptides divided by the number of theoretically observable tryptic peptides for each protein, was used to estimate protein abundance in human spliceosome complex [37]. This index was later converted to exponentially modified PAI (emPAI, the exponential form of PAI minus one) [38]. The emPAI demonstrated its success by determining absolute abundance of 46 proteins in a mouse whole-cell lysate, which had been measured using synthetic peptides. The values of emPAI can be calculated easily with a simple script and do not require additional experimentation in protein identification experiments. It can be routinely used for reporting approximate absolute protein abundance in a large-scale analysis.

Recently, a modified spectral counting strategy termed absolute protein expression (APEX) profiling was developed to measure the absolute protein concentration per cell from the proportionality between the protein abundance and the number of peptides observed [39]. The key to APEX is the introduction of appropriate correction factors that make the fraction of expected number of peptides and the fraction of observed number of peptides proportional to one another. The protein's absolute abundance is indicated by an APEX score, which is calculated from the fraction of observed peptide mass spectra associated with one protein, corrected by the prior estimate of the number of unique peptides

expected from a given protein during a MudPIT experiment. The critical correction factor for each protein (called Oi value) is calculated by using a machine learning classification algorithm to predict the observed tryptic peptides from a given protein based upon peptide length and amino acid composition. APEX successfully determined the abundance of 10 proteins that were spiked in a yeast cell extract with known amounts. The absolute protein abundance of yeast and *E. coli* proteomes analyzed by APEX correlated well with the measurements from other absolute expression measurements such as high-throughput analysis of fusion proteins by western blotting or flow cytometry. The APEX technique has recently been developed as APEX Quantitative Proteomics Tool [40], a free open source Java implementation for the absolute quantification of proteins (http://pfgrc.jcvi.org/).

3. Commercially Available Software for Label-Free Quantitative Proteomics

There has recently been a rapid increase in the development of new bioinformatics tools that aid in automated label-free analysis for comparative LC-MS. The data processing pipelines generally include data normalization, time alignment, peak detection, peak quantification, peak matching, identification, and statistical analysis. Numerous open source and commercial software are available currently. The open source programs include MapQuant, MZmine, MsInspect, OpenMs, MSight, SuperHirn, and PEPPeR [41, 42]. The commercially available software is listed in Table 1. Decyder MS is based on DeCyder 2D Differential Analysis Software. It consists of two main analysis features: peptide detection with the PepDetect module and run-to-run matching with the PepMatch module. PepDetect module provides background subtraction, isotope and charge-state deconvolution, and peak volume calculations using imaging algorithms. This module also provides the option of submitting all or a selected subset of peptides for protein identification by database searching. PepMatch module aligns peptides from different LC-MS runs and detects small quantitative differences between peptides across multiple runs with statistical confidence. Various normalization techniques can be applied to further improve results [18].

SIEVE software employs an algorithm called ChromAlign for chromatographic alignment prior to find differences that are statistically meaningful. The software can determine a P-value for the expression ratio of each differential peak, providing an extra measure of confidence. Peptides that show statistically significant differences can be searched against protein databases to determine peptide and protein

identities. Its prefiltering function reduces the number of spectra that need to be searched, decreases the time spent on identification, and increases the throughput of complex biomarker discovery experiments.

The Rosetta Elucidator system is not only a label-free quantitative software, but also a data management platform to store and manage large volumes of MS data. It also supports labeling analysis such as SILAC and ICAT. Elucidator uses an algorithm called PeakTeller for peak detection, extraction, and quantitation from mass spectrometry data. It uses PeptideTeller and ProteinTeller for verifying correct peptide/protein assignments for all features. The system supports a wide range of MS instruments, database search algorithms, comprehensive visualization and analysis tools [43]. It supports label-free quantification by spectral counting as well.

ProteinLynx Global Server supports label-free quantification by peak intensity. It is also a database searching engine for peptide/protein identification [21, 23].

4. Conclusions

The rapid development of label-free quantitative proteomic techniques has provided fast and low-cost measurement of protein expression levels in complex biological samples. Peak intensity-based comparative LC-MS and spectral count-based LC-MS/MS are the two most commonly used label-free quantification methods. Compared with isotope-labeling methods, label-free experiments need to be more carefully controlled, due to possible error caused by run-to-run variations in performance of LC and MS. However, the development of highly reproducible nano-HPLC separation, high resolution mass spectrometer, and delicate computational tools has greatly improved the reliability and accuracy of label-free, comparative LC-MS. Commercially available data processing software is able to automatically detect, match, and analyze peptides from hundreds of different LC-MS experiments simultaneously, which provides a high-throughput technique for disease-related biomarker discovery. The spectral count-based label-free method positively correlates with isotope-labeling quantification and allows both relative and absolute quantification of protein abundance. These label-free quantitative approaches have provided rigorous, powerful tools for analyzing protein changes in large-scale proteomics studies.

Acknowledgments

References

[1] B. F. Cravatt, G. M. Simon, and J. R. Yates III, "The biological impact of mass-spectrometry-based proteomics," *Nature*, vol. 450, no. 7172, pp. 991–1000, 2007.

[2] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.

[3] H. J. Issaq and T. D. Veenstra, "Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives," *BioTechniques*, vol. 44, no. 5, pp. 697–700, 2008.

[4] T. Kondo, "Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE," *Journal of Biochemistry and Molecular Biology*, vol. 41, no. 9, pp. 626–634, 2008.

[5] A. Motoyama and J. R. Yates III, "Multidimensional LC separations in shotgun proteomics," *Analytical Chemistry*, vol. 80, no. 19, pp. 7187–7193, 2008.

[6] B. Domon and R. Aebersold, "Mass spectrometry and protein analysis," *Science*, vol. 312, no. 5771, pp. 212–217, 2006.

[7] E. I. Chen and J. R. Yates III, "Cancer proteomics by quantitative shotgun proteomics," *Molecular Oncology*, vol. 1, no. 2, pp. 144–159, 2007.

[8] T. D. Veenstra, "Global and targeted quantitative proteomics for biomarker discovery," *Journal of Chromatography B*, vol. 847, no. 1, pp. 3–11, 2007.

[9] V. J. Patel, K. Thalassinos, S. E. Slade, et al., "A comparison of labeling and label-free mass spectrometry-based proteomics approaches," *Journal of Proteome Research*, vol. 8, no. 7, pp. 3752–3759, 2009.

[10] A. H. P. America and J. H. G. Cordewener, "Comparative LC-MS: a landscape of peaks and valleys," *Proteomics*, vol. 8, no. 4, pp. 731–749, 2008.

[11] W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, et al., "Comparison of label-free methods for quantifying human proteins by shotgun proteomics," *Molecular and Cellular Proteomics*, vol. 4, no. 10, pp. 1487–1502, 2005.

[12] R. D. Voyksner and H. Lee, "Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 13, no. 14, pp. 1427–1437, 1999.

[13] D. Chelius and P. V. Bondarenko, "Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry," *Journal of Proteome Research*, vol. 1, no. 4, pp. 317–323, 2002.

[14] P. V. Bondarenko, D. Chelius, and T. A. Shaler, "Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry," *Analytical Chemistry*, vol. 74, no. 18, pp. 4741–4749, 2002.

[15] M. C. Wiener, J. R. Sachs, E. G. Deyanova, and N. A. Yates, "Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures," *Analytical Chemistry*, vol. 76, no. 20, pp. 6085–6096, 2004.

[16] W. Wang, H. Zhou, H. Lin, et al., "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards," *Analytical Chemistry*, vol. 75, no. 18, pp. 4818–4826, 2003.

[17] R. E. Higgs, M. D. Knierman, V. Gelfanova, J. P. Butler, and J. E. Hale, "Comprehensive label-free method for the relative quantification of proteins from biological samples," *Journal of Proteome Research*, vol. 4, no. 4, pp. 1442–1450, 2005.

[18] J. Lengqvist, J. Andrade, Y. Yang, G. Alvelius, R. Lewensohn, and J. Lehtiö, "Robustness and accuracy of high speed LC-MS separations for global peptide quantitation and biomarker discovery," *Journal of Chromatography B*, vol. 877, no. 13, pp. 1306–1316, 2009.

[19] N. Fatima, D. Chelius, B. T. Luke, et al., "Label-free global serum proteomic profiling reveals novel celecoxib-modulated proteins in familial adenomatous polyposis patients," *Cancer Genomics and Proteomics*, vol. 6, no. 1, pp. 41–49, 2009.

[20] S. K. Huang, M. M. Darfler, M. B. Nicholl, et al., "LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis," *PLoS ONE*, vol. 4, no. 2, article e4430, 2009.

[21] Y. Levin, E. Schwarz, L. Wang, F. M. Leweke, and S. Bahn, "Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples," *Journal of Separation Science*, vol. 30, no. 14, pp. 2198–2203, 2007.

[22] J. P. C. Vissers, J. I. Langridge, and J. M. F. G. Aerts, "Analysis and quantification of diagnostic serum markers and protein signaturs for Gaucher disease," *Molecular and Cellular Proteomics*, vol. 6, no. 5, pp. 755–766, 2007.

[23] J. T.-J. Huang, T. McKenna, C. Hughes, F. M. Leweke, E. Schwarz, and S. Bahn, "CSF biomarker discovery using label-free nano-LC-MS based proteomic profiling: technical aspects," *Journal of Separation Science*, vol. 30, no. 2, pp. 214–225, 2007.

[24] M. P. Washburn, D. Wolters, and J. R. Yates III, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology," *Nature Biotechnology*, vol. 19, no. 3, pp. 242–247, 2001.

[25] H. Liu, R. G. Sadygov, and J. R. Yates III, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics," *Analytical Chemistry*, vol. 76, no. 14, pp. 4193–4201, 2004.

[26] B. Zybailov, M. K. Coleman, L. Florens, and M. P. Washburn, "Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling," *Analytical Chemistry*, vol. 77, no. 19, pp. 6218–6224, 2005.

[27] M.-Q. Dong, J. D. Venable, N. Au, et al., "Quantitative mass spectrometry identifies insulin signaling targets in C. elegans," *Science*, vol. 317, no. 5838, pp. 660–663, 2007.

[28] B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn, "Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2339–2347, 2006.

[29] L. Florens, M. J. Carozza, S. K. Swanson, et al., "Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors," *Methods*, vol. 40, no. 4, pp. 303–311, 2006.

[30] A. C. Paoletti, T. J. Parmely, C. Tomomori-Sato, et al., "Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 50, pp. 18928–18933, 2006.

[31] B. Zhang, N. C. VerBerkmoes, M. A. Langston, E. Uberbacher, R. L. Hettich, and N. F. Samatova, "Detecting differential and correlated protein expression in label-free shotgun proteomics," *Journal of Proteome Research*, vol. 5, no. 11, pp. 2909–2918, 2006.

[32] J. X. Pang, N. Ginanni, A. R. Dongre, S. A. Hefta, and G. J. Opiteck, "Biomarker discovery in urine by proteomics," *Journal of Proteome Research*, vol. 1, no. 2, pp. 161–169, 2002.

[33] P. V. Rao, A. P. Reddy, X. Lu, et al., "Proteomic identification of salivary biomarkers of type-2 diabetes," *Journal of Proteome Research*, vol. 8, no. 1, pp. 239–245, 2009.

[34] J. Pan, H.-Q. Chen, Y.-H. Sun, J.-H. Zhang, and X.-Y. Luo, "Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology," *Lung*, vol. 186, no. 4, pp. 255–261, 2008.

[35] J. M. Asara, H. R. Christofk, L. M. Freimark, and L. C. Cantley, "A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen," *Proteomics*, vol. 8, no. 5, pp. 994–999, 2008.

[36] N. T. Seyfried, L. C. Huysentruyt, J. A. Atwood III, Q. Xia, T. N. Seyfried, and R. Orlando, "Up-regulation of NG2 proteoglycan and interferon-induced transmembrane proteins 1 and 3 in mouse astrocytoma: a membrane proteomics approach," *Cancer Letters*, vol. 263, no. 2, pp. 243–252, 2008.

[37] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann, "Large-scale proteomic analysis of the human spliceosome," *Genome Research*, vol. 12, no. 8, pp. 1231–1245, 2002.

[38] Y. Ishihama, Y. Oda, T. Tabata, et al., "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein," *Molecular and Cellular Proteomics*, vol. 4, no. 9, pp. 1265–1272, 2005.

[39] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation," *Nature Biotechnology*, vol. 25, no. 1, pp. 117–124, 2007.

[40] J. C. Braisted, S. Kuntumalla, C. Vogel, et al., "The APEX quantitative proteomics tool: generating protein quantitation estimates from LC-MS/MS proteomics results," *BMC Bioinformatics*, vol. 9, article 529, 2008.

[41] L. N. Mueller, M.-Y. Brusniak, D. R. Mani, and R. Aebersold, "An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data," *Journal of Proteome Research*, vol. 7, no. 1, pp. 51–61, 2008.

[42] U. Rajcevic, S. P. Niclou, and C. R. Jimenez, "Proteomics strategies for target identification and biomarker discovery in cancer," *Frontiers in Bioscience*, vol. 14, pp. 3292–3303, 2009.

[43] H. Neubert, T. P. Bonnert, K. Rumpel, B. T. Hunt, E. S. Henle, and I. T. Lames, "Label-free detection of differential protein expression by LC/MALDI mass spectrometry," *Journal of Proteome Research*, vol. 7, no. 6, pp. 2270–2279, 2008.

Review Article

Challenges for Biomarker Discovery in Body Fluids Using SELDI-TOF-MS

Muriel De Bock,[1] Dominique de Seny,[2] Marie-Alice Meuwis,[3] Jean-Paul Chapelle,[1] Edouard Louis,[4] Michel Malaise,[2] Marie-Paule Merville,[1] and Marianne Fillet[1]

[1] *Laboratory of Clinical Chemistry, GIGA Research, University of Liège, CHU, B36, 4000 Liège 1, Belgium*
[2] *Department of Rheumatology, GIGA Research, University of Liège, CHU, B36, 4000 Liège 1, Belgium*
[3] *Proteomic Platform, GIGA Research, University of Liège, CHU, B36, 4000 Liège 1, Belgium*
[4] *Department of Hepato-Gastroenterology, GIGA Research, University of Liège, CHU, B36, 4000 Liège 1, Belgium*

Correspondence should be addressed to Marianne Fillet, marianne.fillet@ulg.ac.be

Protein profiling using SELDI-TOF-MS has gained over the past few years an increasing interest in the field of biomarker discovery. The technology presents great potential if some parameters, such as sample handling, SELDI settings, and data analysis, are strictly controlled. Practical considerations to set up a robust and sensitive strategy for biomarker discovery are presented. This paper also reviews biological fluids generally available including a description of their peculiar properties and the preanalytical challenges inherent to sample collection and storage. Finally, some new insights for biomarker identification and validation challenges are provided.

1. Introduction

The objective of biomarker discovery is to identify specific protein markers susceptible to improve early diagnosis survey therapeutic outcomes and facilitate the development of novel drug candidates [1, 2]. The methodology relies on differential protein expression profiling. The fundamental approach is based on the assumption that the pathology of concern will affect some physiological processes causing changes in the protein expression levels. Proteins generating similar signals in both sample groups are ignored while significantly up- and downregulated proteins become potential biomarkers. Differential expression profiling requires both a sensitive technology to discern any tiny differences and a high-throughput system in order to process large series of samples required to reach statistical significance. Protein differential display techniques such as two-dimensional gel electrophoresis (2-DE), one- or two-dimensional liquid chromatographic (LC-MS), or surface-enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS) are regarded as the most powerful tools for establishing fingerprint profiles [3–6].

Many reports regarding the application of the SELDI-TOF-MS technology have been published since its introduction in 1993 [7] and its first use for disease detection [8]. One of the key features of SELDI-TOF-MS is its ability to provide rapid protein expression profiles from a variety of biological samples with minimal requirements for purification and separation of proteins prior to mass spectrometry. SELDI-TOF-MS profiling studies revealed that biological fluids contain many proteins with low molecular weight (<15 kDa) not resolved on conventional 2D gels [6, 9].

As can be seen in Figure 1, the SELDI technique consists in surface arrays involving various chromatographic models based on both classic chemistries (normal phase, hydrophobic, cation- and anion-exchange surfaces) and specifically affinity-coated surfaces (immobilized metal affinity capture : IMAC). After the binding phase of the sample to these surfaces, the unbound proteins are washed out while retained molecules are overlaid with an energy-absorbing matrix. In the final step, mass spectra are recorded using a laser for the ionization and a TOF mass spectrometer for its resolving power.
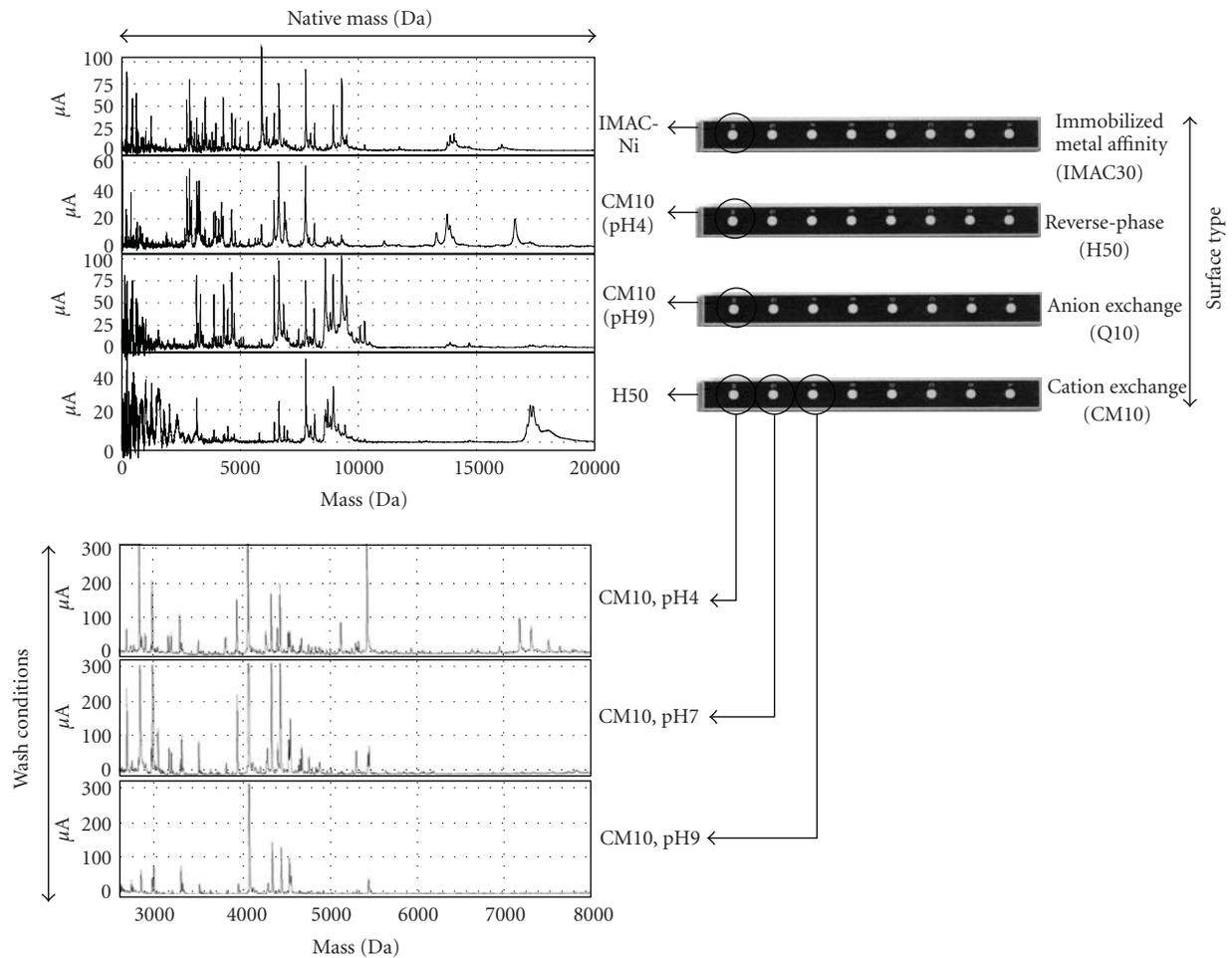
FIGURE 1: Effects of different ProteinChip array surfaces and wash conditions. The combination of ProteinChip array surface types and wash conditions maximize the potential for protein biomarker discovery.

Recent interest in the field has yielded a large number of candidate biomarkers in various diseases [10–35]. However, the small size and poor design of some studies drove validation of these biomarkers quite challenging [36–41].

In the context of clinical proteomic using SELDI-TOF-MS, many recent reviews discussed newly identified disease biomarkers [13, 21, 22, 24, 27, 30, 35, 42–44]. The present review focuses on technical challenges encountered with the SELDI-TOF-MS technology taking into account new insights coming from the last three years. Critical steps that should be undertaken to avoid any bias, to maximize reproducibility and detection sensitivity, with the final aim to find relevant, specific, and robust biomarkers are addressed [45, 46]. For prospective studies, current knowledge on the different biological fluid sources available for SELDI-TOF-MS experiments is described presenting their respective advantages and limitations.

2. Study Design

A successful biomarker research program starts with a careful study design and the preparation of a detailed

protocol. Many manuscripts report encountered problems, emphasizing the importance of Standard Operating Procedures (SOPs), clinical protocols, instrument tuning, and stabilization [37–40, 47–63]. Only critical points will be discussed in this review.

In the early phase of biomarker discovery, the clinical question addressed has to be defined in the disease(s) context collecting adequate control samples. Indeed, it can be criticized that in many published studies, patients were compared to healthy subjects rather than to patients presenting similar diseases or clinical signs.

Experimental workflow and technologies have to be selected with great care. The avoidance of bias is not trivial and must be addressed throughout the whole study, from its design to the data analysis and interpretation (cf. Figure 2). Current proteomics and genomics technologies are extremely sensitive and can detect very small changes in expression levels. Some of these changes may arise from biological differences related to disease or pharmacological treatment. They could also result from the heterogeneity of the patient panels tested across multiples sites, the inherent biological complexity, and the diversity of sample
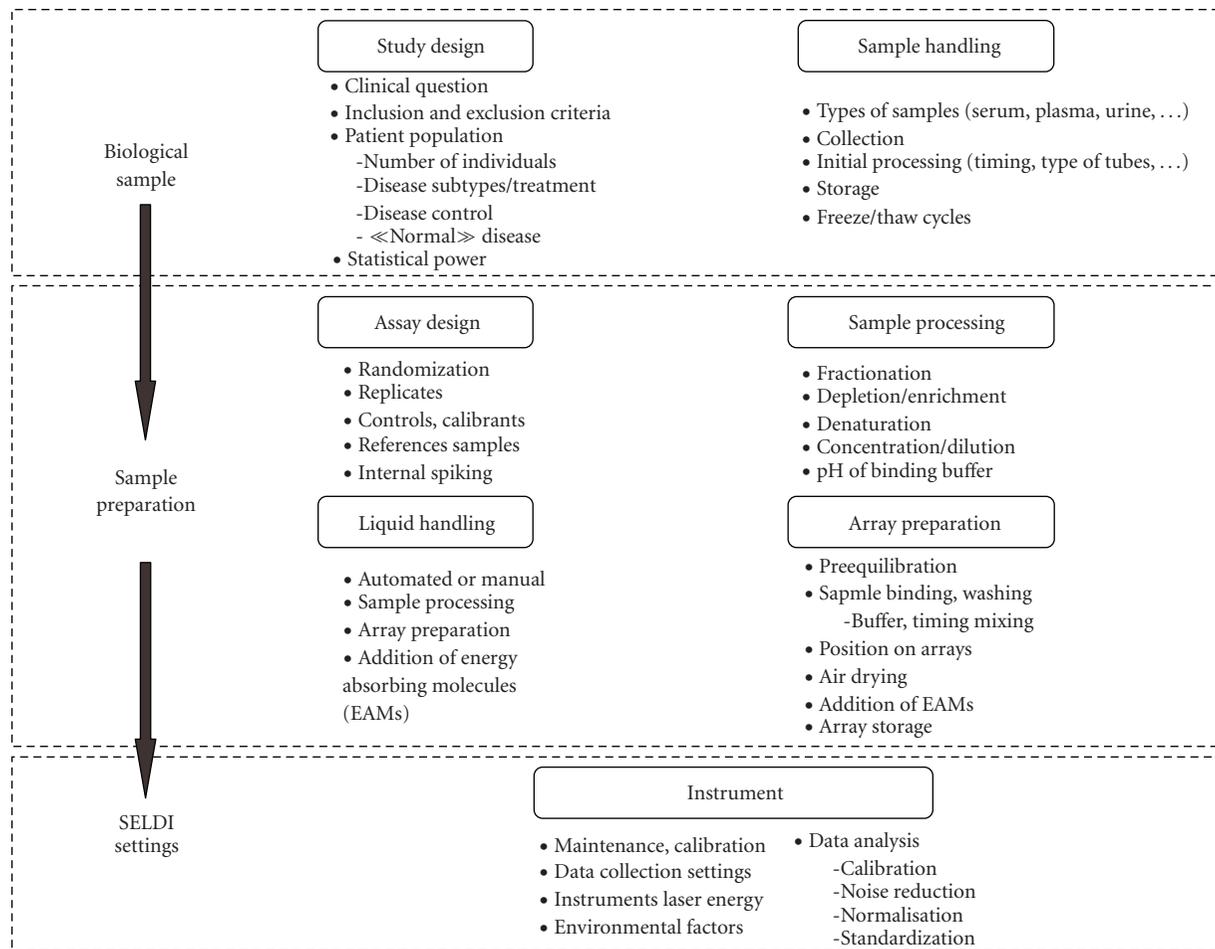
FIGURE 2: Experimental variables that can affect proteomics data. Most of the steps shown are involved in all proteomics workflows, but SELDI technology performs many of them on a single platform.

types. Small differences in sample collection, processing, and analytical techniques could have some impact on the outcomes of the study. As a consequence, clinical data may be site-, study-, population-, or sample-dependent, without any actual clinical relevance [53, 62, 64–66]. The key factor for maximizing reproducibility in biomarker research is to identify and minimize all potential sources of preanalytical and analytical bias [53, 55] (Table 1). Adherence to strict guidelines and SOPs is critical to reach the highest operating standards for data quality and reproducibility [37, 38, 47, 50, 51, 53, 55, 62, 63, 67]. SOPs also facilitate the validation of biomarkers by other groups using different sets of samples.

3. Sample Handling and Preparation

Besides the instrumentation and the methodologies related to chromatography-mass spectrometry analysis, the nature, quality, and number of clinical samples to process are key elements to be considered for any proteomic approaches.

3.1. Selection of Body Fluid. In order to provide positive answer to any precise clinical question, the investigator has to make a choice among the most relevant biological target samples (body fluid, tissue, etc.). Many criteria must be considered at this level, that is, availability, easiness of collection, stability, composition, proximity with disease location, patient discomfort, ethics, and so forth.

For biomarker discovery using SELDI-TOF-MS, a great variety of biological sample types can be used; major applications concern biofluids (plasma, serum, urine, saliva, cerebrospinal liquid, bronchoalveolar wash out, nipple aspirate fluid, tears, amniotic fluid) [1, 10–14, 18, 19, 21, 22, 24, 28, 42, 48, 49, 68–77], and tissues or cell extracts [78, 79]. Among them, serum, plasma, urine, and saliva are the most popular. Many variables related to collection, storage, and conditions for sample preparations have to be carefully considered. These parameters have been commented in literature according to the nature of the biofluids [48, 49, 54, 57, 61, 67, 73, 74, 80–84]. In this section of the review, consensus opinions derived from these recent studies are provided. Advantages and drawbacks of the most popular

TABLE 1: Factors that impact preanalytical and analytical bias.

Preanalytical bias	
Patients information	Age, gender, ethnicity
	Disease subtype and/or severity
	Medical background
	Health background
	Smoking status, alcohol intake, diet, other risk factors
	Drug treatments
	Patient position (seated/standing/lying), daily moment of collection
	Type of control (healthy or disease)
	Location of sample collection (single or multisite)
	Study inclusion and exclusion criteria
Sample characteristics	Number of individuals
	Type (blood, serum, plasma, urine, cerebrospinal fluid, cell lysate, etc.)
	Source (banked or prospectively collected)
Sample-handling procedures	Collection protocols (initial processing, procedure, timing, type of anticoagulant, type of tubes, number of sites, etc.)
	Storage procedures (time, aliquoting, storage materials, temperature, freeze-thaw cycles, etc.)
Analytical bias	
Sample-Processing procedures	Fractionation and depletion methods
	Processing steps (denaturation, buffer components, delipidation, etc.)
	Liquid handling methods (automated or manual, technique, equipment, etc.)
Experimental protocols	Array types
	Sample pH and dilution factor
	Quantity of sample loading and position on arrays
	Sample binding, washing and drying procedures
	Matrix addition (type and method)
	Instruments settings
	Number of instruments, locations
	Environmental factors (temperature, humidity percentage)
Data analysis methods	Spectrum processing (baseline subtraction, normalization, alignment, noise reduction, etc.)
	Peak labelling
	Feature selection, statistical analysis
	Classification approaches

fluids are presented in Table 2 as well as formulated recommendations. When reviewing a series of studies [47–49, 51, 54, 57, 58, 60, 61, 67, 68, 73, 81–83, 85] general guidelines can be forwarded. Optimal serum clotting arises after 60 minutes at room temperature. After clot formation, samples can be transported or stored on wet ice for 3 hours before centrifugation. Aliquots must be prepared and stored at $-80°C$. For plasma collection, anticoagulant EDTA is preferred and its processing should be realized as soon as possible after sampling, ideally within the first hour. Although storage at low temperature promotes peptides and proteins stability, one should not recommend storing plasma samples at $4°C$ due to the cold activation of platelets. Prior any freezing, plasma can be depleted in platelets by using a filtration step; aliquots are then frozen at $-80°C$.

According to the "HUPO PPP Specimens Committee" recommendations, plasma appears preferable to serum because it contains less peptides of degradation and consequently presents less variability [57, 81, 86]. In order to avoid the presence of platelet related peptides, the authors also recommend to use platelet-poor plasma obtained by centrifugation followed by a filtration step. However, the choice of serum could be justified when studying diseases related to coagulation abnormalities. Furthermore, it is often more available in sample banks for retrospective studies.

A controversial parameter is the addition of protease inhibitors (PIs) to the samples. Some authors found that the addition of a PI cocktail induces significant differences in protein profiles when compared to crude samples [58, 83]. Whenever directly introduced during phlebotomy, PI allows

TABLE 2: Advantages and limitations of body fluids particularly useful for biomarker discovery.

Body fluids	Advantages	Limitations	Recommendations
Serum	(i) established sample banks often composed of serum aliquots (retrospective studies), (ii) proteins and peptides that "survive" to the clotting procedure exhibit a stability that can be exploited in routine clinical applications.	(i) presence of various products derived from coagulation cascade, (ii) biomarker with poor stability during coagulation process will not be detected in serum, (iii) possible influence of the disease on coagulation process.	(i) use standardized collection protocol, (ii) keep sample during 1 hour at RT to allow clotting process before centrifugation, (iii) preserve on ice after clotting. Aliquoting and freezing ($-80°C$) cannot be done immediately.
Plasma	(i) more rapidly processed than serum (interesting for emergency diagnosis), (ii) larger final volume of fluid after processing than with serum, (iii) more stable than serum due to the inhibition of coagulation cascade.	(i) interference with chip surface (i.e. heparin tube), (ii) sample dilution in citrate tube, (iii) possible interference of EDTA with protein binding on IMAC surface, (iv) SELDI-TOF spectra less rich in peaks number and intensity than serum.	(i) use standardized collection protocol, (ii) carefully choose the type of anticoagulants (EDTA tubes are preferable), (iii) use platelet-poor plasma, (iv) centrifuge, aliquot and freeze ($-80°C$) as soon as possible. If not possible, keep at RT to avoid cold platelet activation.
Dry blood	(i) medical staff not needed for collection, (ii) low blood volume necessary, (iii) easy storage and transport.	(i) elution step to recover sample from filter paper.	(i) keep dry specimens at RT for 3–4 hours in horizontal position, (ii) store at $-20°C$.
Saliva	(i) easy and noninvasive sampling, (ii) medical staff not needed for collection,	(i) low volume collected, (ii) presence of many proteases and unspecific materials such as food residues or microorganisms, (iii) level of certain plasma proteins are not reflected in saliva.	(i) always collect with the same method (stimulated or not) and at the same moment of the day, (ii) centrifuge to remove insoluble material, aliquot and freeze at $-80°C$.
Urine	(i) easy and noninvasive collection, (ii) medical staff not needed for collection, (iii) obtained in large volume.	(i) fluctuation of protein concentration overtime and according to renal integrity, (ii) presence of salts and proteins in low concentration.	(i) use standardized collection protocol, (ii) concentrate the samples, (iii) centrifuge, aliquot and freeze at $-80°C$, (iv) normalization with creatinine content.

fluid stabilization for at least 2 hours at room temperature by reducing proteolysis damages. However, PI presents some additional drawbacks such as the presence of highly concentrated components in the cocktail which can compete later on for protein array interactions.

Another important factor to decrease the risk of variation, bias, and errors is the communication between researchers and medical staff. One generally considers that 70% of the errors are due to human intervention (mostly due to communication problems) while only 30% appear instrumental related errors [87]. The mode of specimen collection (veni-puncture or arterial puncture), the site of collection, the position of the patient, or the tourniquet technique can influence the concentration of certain blood constituents [58]. Hemolysis also causes significant changes in blood proteome specimens [67]. It is generally advised to discard those kinds of samples, but when the disease studied involves spontaneous hemolysis, this cannot reasonably be done.

Less commonly used, filter papers were also described to collect blood [68]. This mode of collection has the advantage that only few drops of blood are needed (particularly interesting for neonatal and repeated screening). Moreover, it does not require specific medical support for sampling, which could be promising for multiple collects realized by the patient at home, in the perspective of a treatment follow-up, for instance. Stability and reproducibility of this collection mode remain to be studied.

Saliva and urine have more recently presented an interest in biomarker discovery. Their collection is simple, noninvasive, and cheap and can be easily repeated. However, like blood specimen, such factors have to be taken into consideration to improve reproducibility of sample collection. Saliva protein composition varies with circadian rhythm, diet, age, gender, and physiological status [86]. It is also affected by the method of sample collection (stimulated versus nonstimulated saliva production) [60]. Food ingestion increases the proteolysis activity and then collection before lunch rather than after is recommended [73]. The addition of PI can reduce but not completely eliminate the impact of the proteolysis [60]. It will stabilize, qualitatively and quantitatively, the saliva proteome for up to 48 hours [73]. Regarding storage conditions, it is preferred to store the saliva specimens at -80°C rather than at -20°C where the preservation of the protein content could not be guaranteed for more than 1 month. Interestingly, repeated freeze-thaw cycles (4/5) do not seem to significantly alter saliva protein profile [74].

Urine has the advantage that it can be obtained in large volume. It is mainly an aqueous solution (95% of water) of waste electrolytes and metabolites, organic components (urea, uric acid), and proteins at low concentrations in healthy individuals (150 mg/day). Urine proteome variation depends mainly on plasma composition due to its role as blood content regulator and on the integrity of the glomerular filtration step leading to a large intra and intersubjects variability. Protein and salt concentrations can vary along the day for a same subject (first void compared to midstream urine samples) [80, 88]. Progressive degradation of urine

proteome due to proteolytic activity can be prevented by PI addition only up to 2 hours of storage [54]. As already mentioned for blood and saliva, up to 5 freeze/thaw cycles do not significantly affect urine proteome profile. Storage at -80°C is still requested.

Other fluids such cerebrospinal fluid, nipple aspirates, tears, synovial fluid, bronchoalveolar lavage, follicular, and amniotic fluids have already been explored by SELDI-TOF-MS [15, 20, 33, 34, 69, 75, 89, 90]. These fluids are generally used to study well-localized diseases. Despite the presence in such fluids of some plasma proteins, their implication to study systemic diseases is not recommended and difficult to apply in routine diagnosis due to risk and discomfort related to collection.

3.2. Sample Processing. One of the most challenging aspects in studying body fluids protein profiles remains the detection of the deep proteome [91]. The protein concentration dynamic range detectable by means of MALDI-TOF or SELDI-TOF-MS is about 2 orders of magnitude, whereas the range in blood reaches about 10 orders of magnitude [91, 92]. As protein binding onto chromatographic surface depends on its affinity, its concentration, but also on the surface binding capacity, one can imagine that the competition between different proteins for binding sites is very complex. A highly abundant protein with low affinity for the chip surface and a low abundant protein with high affinity may give similar peak intensities in the final SELDI mass spectrum. Furthermore, protein steric hindrance can also affect the SELDI profiles.

Several fractionation procedures are now available to decrease the sample protein concentration dynamic range [85, 93–102].

A major inconvenience for sample fractionation is the resulting low sample throughput capacity, due to a significant increase of the duration of analysis and to a risk of poor reproducibility affecting data treatment. Use of automatized technologies can improve the reproducibility and decrease the total analysis time. Additionally the same proteins can be presented in different fractions challenging the comparison of their abundance between samples.

Several methods have been proposed for fractionation such as centrifugal ultrafiltration, precipitation by organic solvents, electrophoresis, chromatography (on-column or on-magnetic beads), or subcellular localization. The choice will be made based on the nature of the sample to be analysed and the protein properties (molecular weight, localization, abundance, etc.). All these sample preparation methods have already been discussed by other reviewers [85, 94, 101, 103, 104]. Recently, with the growing interest in studying posttranslational modifications new methodologies set up to isolate rare amino acid-containing peptides (cys, met, trp, his) or PTM peptides (phosphopeptides, glycopeptides) have been developed [25]. One of the most widely used approach for highly abundant proteins removal in serum and plasma is their depletion using antibodies. Despite the depletion of the nine most abundant proteins from serum or plasma samples, overall published results were quite disappointing [105]. This

sensitivity problem is most probably inherent to the too low concentration of the peptidome constituents. Moreover, some of the abundant proteins act as carrier explaining the codepletion of almost 3000 species (peptides and proteins) as observed by several groups [106].

A new fractionation approach has been recently developed by Righetti and Boschetti [107]. It implied a solid-phase combinatorial library of hexapeptides on which millions of copies of a unique ligand are graft on a bead. This technique, enabling the dilution of abundant protein by rapid saturation of its ligand, concentrates components of the deep proteome which could not reach saturation. This method presents the advantage to reduce the dynamic range between the most and less abundant proteins and peptides. It has also been showed that despite compression of the dynamic range, this technology used for differential studies was only applicable for proteins or peptides which do not reach saturation (low and medium abundance proteins) [107]. Many studies conducted on different types of samples report good reproducibility and important gain in the number of low abundant species by comparison with analyses performed on corresponding crude samples [96, 98, 99, 108–111], which make this approach very promising to investigate the deep proteome.

4. SELDI Settings

In order to highlight candidate protein biomarkers, several chromatographic surfaces must be screened. The choice of the protein chip array chemistry and the nature of the matrix depend on whether the application requires general profiling or requires a specific protein assay. Different array types and binding conditions may generate complementary protein profiles for the same sample [7]. The use of relevant quality controls (QCs) is highly recommended and even mandatory in such applications [37–40, 47–63]. QCs should be well-characterized pools of samples processed alongside the experimental samples in order to monitor instrument performances, optimize mass spectrometry settings (laser energy, etc.), compare target protein profiles to those of historical reference samples, and to calculate coefficient of variation for peak intensities as a measure of reproducibility.

It is important to point that the resolution and mass accuracy provided by this kind of instrument are rather low compared to high-resolution mass spectrometers (i.e., Q-TOF, FT-MS, etc.). Using SELDI-TOF-MS, one could not expect to accurately determined m/z values or peak intensities on complex mixtures. Indeed, low resolution causes peaks overlap making abundance and mass assignment difficult. This means that only large differences in peak intensities are to be considered and that peaks of interest have to be identified with more accurate mass spectrometers. Beside those instrumental weaknesses, on the contrary to other mass spectrometers, SELDI-TOF-MS can be used for high throughput analysis.

During SELDI settings, numerous sources of spectra variability have to be taken into account.

Several events, such as matrix crystallization, ion suppression, and in-source decay occurring during mass spectra acquisition strongly influence the peak intensities. These are commented in more details below.

4.1. Matrix Crystallization. Differences in reagents, handling of material, room temperature, and level of humidity may all influence the (co)crystallization step of matrix molecule with sample causing interday fluctuation. The structure and nature of the target surface may also affect peak intensities. These parameters must be highly controlled and standardized for each study protocol. During the crystallization process, a competition phenomenon can occur between proteins for crystal inclusion. Easily embedded proteins will be present at higher concentrations in the matrix and consequently more efficiently desorbed and ionized [47]. To improve sample-to-sample reproducibility of MALDI ion yield and to increase the precision of peptide quantification, some authors use nitrocellulose in order to improve the homogeneity of the matrix/analyte crystallization [55, 112, 113]. This operation might also be helpful for SELDI-TOF-MS measurements.

4.2. Ion Suppression. Depending on sample composition, ion suppression is another factor that significantly contributes to the variability observed in SELDI-TOF-MS spectra [47, 50, 55]. Indeed, during ionization, analytes compete for protons that are transferred from matrix molecules. If a protonated analyte collides with an unprotonated one which has higher gas-phase basicity, it may pass its proton to the collision partner. Therefore, the presence of an analyte may reduce the signal intensity of another. This phenomenon is called "ion suppression effect." In a complex protein mixture like serum, where highly abundant proteins constitute a large proportion of the total protein content, it is possible that such peaks override signals from low abundant peptides. This phenomenon, obviously difficult to prevent in complex samples, would be more easily controlled on mixtures issued from fractionation.

4.3. In-Source-Decay. Another source of variations is the fragmentation of proteins or peptides during mass spectrometric process. Fragmentation occurring before the first field-free region is called in-source decay (ISD); it is responsible for consecutive series of ions [114].

Ekblad et al. showed that ISD generates quite additional spectral peaks in the spectrum of proteins contained in serum samples when compared with the data collected for pure reference proteins [114]. One obviously creates ISD favourable conditions when optimizing the analytical conditions by maximizing the total peak count, particularly when using a high laser beam which would increase the thermal ions energy and consequently the number of collisions between ions. Hopefully, in-source fragmentation remains quite limited [114]. Dijkstra et al. developed a method which deconvolutes the spectrum by appropriately associating peaks belonging to the same protein [50]. To take benefit

of this procedure, highly efficient sample fractionation is recommended.

4.4. Miscellaneous. Other phenomena susceptible to affect SELDI spectra must be considered. Common mechanisms accounting for the arising of multiple peaks in mass spectra include, for example, the formation of salt adducts and multiply charged ions [50]. Chemical reactions using energy from the laser may take place between sample protein molecules, matrix molecules, or molecules from the washing buffers generating intermolecular complexes known as "ions cluster" [50]. The formation of these complexes increases the number of spectrum peaks causing artefacts (i.e., satellite peak at +206 Da corresponds to a SPA adduct). Moreover, the performance of the SELDI-TOF-MS may change over time due to possible fluctuation in the laser intensity and/or detector sensitivity.

All these difficulties can be addressed only by substantial reduction in sample complexity and the application of a rigorous standardization program of the entire analytical process. This involves optimized acquisition protocols (i.e., avoiding too high laser intensity), a fully operational and calibrated instrument and the use of suitable QC samples, similar in nature and complexity to the studied samples.

5. Data Analysis

5.1. Spectrum Processing. Another important methodological source of artefacts is the data analysis of protein profiles. The data preprocessing (calibration, baseline correction, normalization, peak detection, and peak alignment) represents a key step for SELDI analysis [115–117].

Spectra are generally normalized in order to equalize or minimize differential effect due to external variation [59, 115, 116, 118, 119]. The widely used total ion current (TIC) gives a clear indication of the impacts of technical variables such as laser and detector performances, matrix application, and sample amounts. TIC normalisation relies on the assumption that the technical parameters are mostly responsible for the largest differences observed between samples. But Cairns et al. showed that TIC may also potentially remove some pertinent biological information [115]. They suggest to examine whether normalisation factors vary systematically between study groups and they recommend to specify the applied methodology (local or global normalisation, matrix signal excluded or not). The ideal normalisation procedure would be to resort to some internal spiking method.

5.2. Classification Approaches. One important aspect in SELDI-TOF-MS data analysis is to avoid false discovery of protein peaks, for which the discriminative power results from random variation. A general criticism concerns the use of inadequate algorithms for data analysis and the problem issued from over-adjustment in combination of high-dimensional data with a low number of cases. Those artefacts could be prevented by analysing a sufficient number of samples, by resorting to overfitting-resistant algorithms, by an appropriate validation of the resultant model, and

by using optimal spectra processing techniques (calibration, exclusion of spectral regions affected by high noise, peak alignment, and normalization). Two others remarks can be formulated from literature reports: (1) multiple biomarkers have generally a better predictive value than individual markers, and (2) positive-predictive values of peptide patterns are often insufficient to be recognized as early markers when they concern low-frequency diseases in the population [38, 53].

The most commonly used bioinformatics approaches are decision tree-based ones and support vector machines [120, 121]. Authors generally emphasized on the need for validated model selection using cross validation loop and permutation testing to develop generalized classifier able to correctly predict classification of new samples [122].

6. Biomarkers: From Identification to Clinical Application

Identification of candidate biomarkers, while not strictly necessary for diagnostic purpose, can be regarded as extremely satisfying in helping to data interpretation and better understanding the disease. As often criticized, the SELDI-TOF-MS technology does not provide peptide/protein identification. In order to succeed in the identification by sequencing (Q-TOF, TOF-TOF, ion-trap, etc.) or peptide fingerprinting (MALDI-TOF), enrichment and purification of the biomarker of interest is often needed, which is laborious and time consuming. To solve in part this weakness, new ProteinChip interface coupled to tandem mass spectrometer was recently developed allowing direct sequencing of peptides <6000 Da [124]. In all cases, identifications must be corroborated using antibody-based detection (i.e., Western blot or ELISA) or antibody pull-down with subsequent detection by SELDI-TOF-MS.

It should be noted that the concentration range of widely used biomarkers in plasma samples is remarkably wide and differ from the high milligram until low nanogram per liter range. For example, serum albumin, within a normal concentration range of 35–50 mg/mL, is measured as an indication of severe liver disease [125] or malnutrition [126], whereas IL-6 normally varies in a range of 0–5 pg/mL, is measured as a sensitive indicator of inflammation or infection [127].

Until now, most of the markers identified after an SELDI-TOF-MS study could not yet be considered as very specific of a given disease but they are rather representative of disease's consequences like inflammation or immune response. The most frequently identified proteins so far are haptoglobulin, transthyretin, apolipoproteins, serum amyloid, or complement factors present at μg/mL to mg/mL [13, 19, 23–25, 28, 38]. Although individual acute-phase reactions proteins are not satisfactory diagnosis biomarkers, their combined use with other serum biomarkers may enable more sensitive and specific diagnosis (cf., Figure 3). This phenomenon has recently been termed "host response protein amplification cascade" [122]. Acute-phase proteins could also be directly produced by the disease tissue.
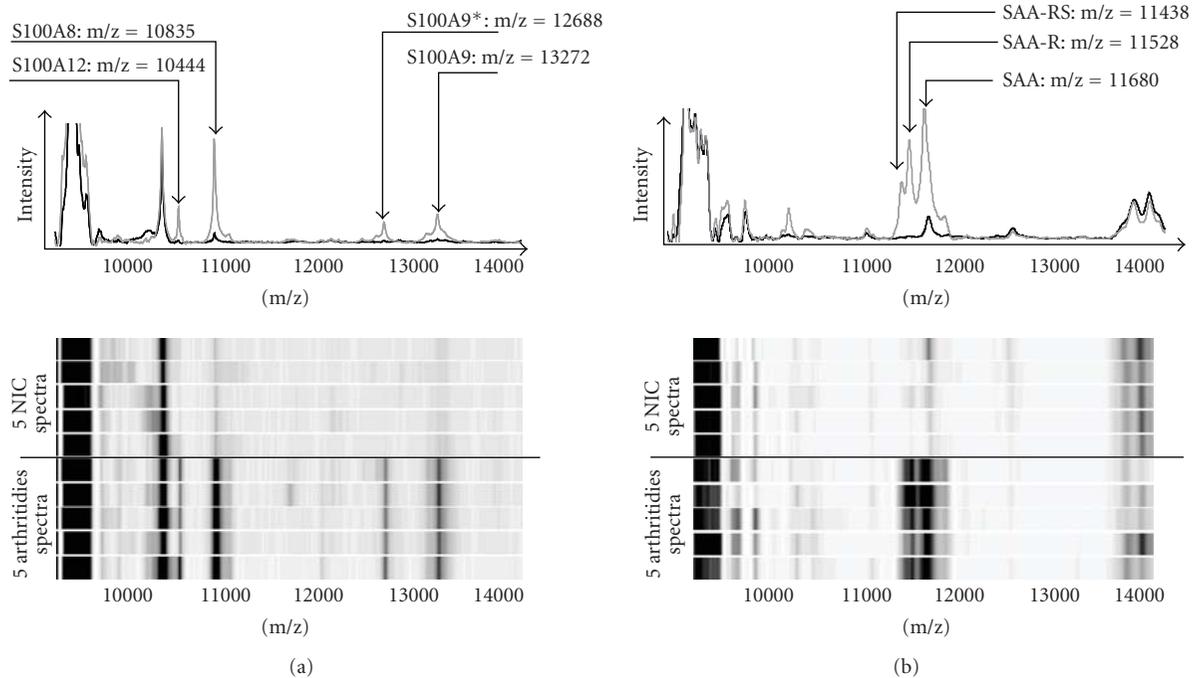
FIGURE 3: Protein mass spectra collected on CM10 and IMAC-Cu^{2+} ProteinChip arrays with serum samples provided by five patients with arthritides (including rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis) and five noninflammatory controls (NIC) (including osteoarthritis). (a) The inflammatory-related proteins S100A8, S100A12, S100A9, and one of its variant S100A9* are arthritis biomarkers detected on CM10 arrays. (b) On IMAC-Cu^{2+} ProteinChip arrays, SAA and its 2 variants (SAA-R and SAA-RS) are illustrated, reproduced from [19].

The moderate specificity of SELDI-discovered biomarkers could be explained by its low sensibility. To date, SELDI-TOF-MS has not yet identified any protein marker present at ng/mL level. This probably indicates that the lowest detect limit of this technology is around μg/mL as considered by Diamandis [128]. To overcome this limited detection sensitivity, the serum (or plasma) proteins can be fractionated (cf., Section 3.2) before SELDI-TOF-MS analysis. Fractions could then be loaded on different arrays using complementary binding conditions.

Moreover, the decisive advantage of the mass spectrometry technologies is the capacity to detect protein variants, protein fragments, and posttranscriptional modifications (PTMs), which is usually not possible with affinity-based technologies. It is now recognized that those components may be disease-specific and can be considered as potential biomarkers (i.e., modified transthyretin forms in ovarian cancer in Figure 4 and in familial amyloidotic polyneuropathy) [123, 129, 130].

In the last two years, lots of applications using SELDI-TOF-MS were published for diagnostic of cancers [25, 42, 44, 131], especially breast [10, 17], prostate [21, 132, 133], and colorectal cancer [24]. Other recent papers concerning infectious diseases [22], neurodegenerative disorders [35], renal diseases [26, 134], and chronic inflammatory diseases [19, 135] also demonstrated the great potential of the technique.
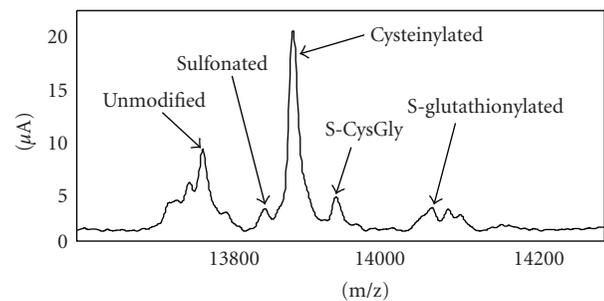
FIGURE 4: Modified transthyretin forms observed in ovarian cancer sample, adapted from [123].

SELDI-TOF-MS technology was also used to predict response to therapy, particularly in cancers. Röcken and Whelan described in detail the use of SELDI-TOF-MS to not only predict responses to cancer therapy but also demonstrate its interest in the follow-up of metastasis disease progression and in the development of drug resistance [44, 136]. Recently platelet factor 4 (PF4) appeared to be a biomarker for Infliximab nonresponse in Crohn's disease and rheumatoid arthritis [29, 137].

For most of these studies, a validation phase should assess the validity of the described potential biomarkers against a larger and more heterogeneous population of patients. The

robustness of the candidate markers has to be tested against a level of biological variability that more accurately reflects the variability in the target population.

Unfortunately, several groups failed to validate the biomarker discovered in their pilot study, such as McLerran et al. [39, 40], and others [36, 66]. McLerran et al. described preanalytical bias. They concluded that their first study samples most likely had biases in the sample selection. Another validation performed by Engwegen et al. using distinct patient populations confirmed that SAA peak clusters are associated to renal cell carcinoma. However, some other markers could not be validated [36]. Such examples demonstrate the importance to strictly control parameters such as storage, clotting, time of analysis, instrument performances, sample selection, and statistical classification method.

The urgent need for SOPs in clinical proteomics research is therefore absolutely mandatory reflecting a growing trend in the field [19, 53, 62, 63]. Interaction between researchers, clinicians, and statisticians is also a key element for the success.

Altogether, these applications of SELDI-TOF-MS technology illustrate its capability for discrimination and follow-up of a multitude of diseases using different body fluids as well as certain therapeutic response prediction. It is worth mentioning that FDA approved recently the first diagnostic tool (named OVA1) issued from SELDI proteomic researches. It is made of the combination of 5 markers for ovarian cancer diagnostic.

7. Concluding Remarks

Taking into account herein and previously described recommendations, SELDI-TOF-MS offers very exciting opportunities to discover not only diagnostic but also prognostic and mechanistic markers for a number of major diseases.

To face the general criticism, standardized procedures and recommendations to minimize bias are now followed by most of the users. However, some challenges still remain, as for all other proteomic approaches, due in part to the complexity and the wide dynamic range of the samples. Sample fractionation and/or enrichment procedure, such as peptide ligand affinity beads, will certainly be the solution to visualise the deep proteome. In addition, improvements in mass spectrometry instrumental performances could be expected (higher resolution, reducing adduct formation, and ion suppression), contributing further to more reliable and faster biomarkers discovery.

Abbreviations

2-DE:	Two-dimensional electrophoresis
CSF:	Cerebrospinal fluid
EDTA:	Ethylenediaminetetraacetic acid
ELISA:	Enzyme-Linked Immunosorbent Assay
HUPO PPP Specimens Committee:	Plasma Proteome Project
IMAC:	Immobilized Metal Affinity Capture
ISD:	In-Source Decay
LC-MS:	Liquid Chromatography—Mass Spectrometry
MALDI-TOF:	Matrix-Assisted Laser Desorption/Ionisation—Time-Of-Flight
PI:	Protease inhibitor
PTM:	Post-translational modifications
PSA:	Prostate specific antigen
QC:	Quality control
SELDI-TOF-MS:	Surface-Enhanced Laser Desorption/Ionisation—Time-Of-Flight—Mass Spectrometry
SOP:	Standard Operating Procedures
SPA:	Sinapic acid
SVM:	Support Vector Machine
TIC:	Total Ion Current.

Acknowledgment

References

[1] L. A. Liotta, M. Ferrari, and E. Petricoin, "Clinical proteomics: written in blood," *Nature*, vol. 425, no. 6961, p. 905, 2003.

[2] A. D. Westont and L. Hood, "Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine," *Journal of Proteome Research*, vol. 3, no. 2, pp. 179–196, 2004.

[3] A. Gorg, W. Weiss, and M. J. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, 2004.

[4] G. Mitulovic, C. Stingl, M. Smoluch, et al., "Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests," *Proteomics*, vol. 4, no. 9, pp. 2545–2557, 2004.

[5] R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads, and T. D. Veenstra, "Characterization of the low molecular weight human serum proteome," *Molecular & Cellular Proteomics*, vol. 2, no. 10, pp. 1096–1103, 2003.

[6] G. L. Wright Jr., "SELDI proteinchip MS: a platform for biomarker discovery and cancer diagnosis," *Expert Review of Molecular Diagnostics*, vol. 2, no. 6, pp. 549–563, 2002.

[7] T. Hutchens and T. Yip, "New desorption strategies for the mass spectrometric analysis of macromolecules," *Rapid Commun Mass Spectrom*, vol. 7, pp. 576–580, 1993.

[8] G. L. Wright Jr., L. H. Cazares, S.-M. Leung, et al., "ProteinChip® surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate

cancer biomarkers in complex protein mixtures," *Prostate Cancer and Prostatic Diseases*, vol. 2, no. 5-6, pp. 264–276, 1999.

[9] N. Tang, P. Tornatore, and S. R. Weinberger, "Current developments in SELDI affinity technology," *Mass Spectrometry Reviews*, vol. 23, no. 1, pp. 34–44, 2004.

[10] M. Abramovitz and B. Leyland-Jones, "A systems approach to clinical oncology: focus on breast cancer," *Proteome Science*, vol. 4, p. 5, 2006.

[11] F. Bertucci, D. Birnbaum, and A. Goncalves, "Proteomics of breast cancer: principles and potential clinical applications," *Molecular and Cellular Proteomics*, vol. 5, no. 10, pp. 1772–1786, 2006.

[12] F. Bertucci and A. Goncalves, "Clinical proteomics and breast cancer: strategies for diagnostic and therapeutic biomarker discovery," *Future Oncology*, vol. 4, no. 2, pp. 271–287, 2008.

[13] J. A. P. Bons, M. P. van Dieijen-Visser, and W. K. W. H. Wodzig, "Clinical proteomics in chronic inflammatory diseases: a review," *Proteomics Clinical Applications*, vol. 1, no. 9, pp. 1123–1133, 2007.

[14] C. S. Buhimschi, V. Bhandari, B. D. Hamar, et al., "Proteomic profiling of the amniotic fluid to detect inflammation, infection, and neonatal sepsis," *PLoS Medicine*, vol. 4, no. 1, article e18, 2007.

[15] I. A. Buhimschi, E. Zambrano, C. M. Pettker, et al., "Using proteomic analysis of the human amniotic fluid to identify histologic chorioamnionitis," *Obstetrics and Gynecology*, vol. 111, no. 2, part 1, pp. 403–412, 2008.

[16] I. A. Buhimschi, G. Zhao, V. A. Rosenberg, S. Abdel-Razeq, S. Thung, and C. S. Buhimschi, "Multidimensional proteomics analysis of amniotic fluid to provide insight into the mechanisms of idiopathic preterm birth," *PLoS ONE*, vol. 3, no. 4, article e2049, 2008.

[17] C. H. Clarke, J. A. Buckley, and E. T. Fung, "SELDI-TOF-MS proteomics of breast cancer," *Clinical Chemistry and Laboratory Medicine*, vol. 43, no. 12, pp. 1314–1320, 2005.

[18] D. de Seny, M. Fillet, M.-A. Meuwis, et al., "Discovery of new rheumatoid arthritis biomarkers using the surface-enhanced laser desorption/ionization time-of-flight mass spectrometry proteinchip approach," *Arthritis and Rheumatism*, vol. 52, no. 12, pp. 3801–3812, 2005.

[19] D. de Seny, M. Fillet, C. Ribbens, et al., "Monomeric calgranulins measured by SELDI-TOF mass spectrometry and calprotectin measured by ELISA as biomarkers in arthritis," *Clinical Chemistry*, vol. 54, no. 6, pp. 1066–1075, 2008.

[20] J. He, J. Gornbein, D. Shen, et al., "Detection of breast cancer biomarkers in nipple aspirate fluid by SELDI-TOF and their identification by combined liquid chromatography-tandem mass spectrometry," *International Journal of Oncology*, vol. 30, no. 1, pp. 145–154, 2007.

[21] M. Hellstrom, H. Lexander, B. Franzen, and L. Egevad, "Proteomics in prostate cancer research," *Analytical and Quantitative Cytology and Histology*, vol. 29, no. 1, pp. 32–40, 2007.

[22] A. Hodgetts, M. Levin, J. S. Kroll, and P. R. Langford, "Biomarker discovery in infectious diseases using SELDI," *Future Microbiology*, vol. 2, no. 1, pp. 35–49, 2007.

[23] J. T.-J. Huang, L. Wang, S. Prabakaran, et al., "Independent protein-profiling studies show a decrease in apolipoprotein A1 levels in schizophrenia CSF, brain and peripheral tissues," *Molecular Psychiatry*, vol. 13, no. 12, pp. 1118–1128, 2008.

[24] S. Hundt, U. Haug, and H. Brenner, "Blood markers for early detection of colorectal cancer: a systematic review," *Cancer Epidemiology Biomarkers and Prevention*, vol. 16, no. 10, pp. 1935–1953, 2007.

[25] K. K. Jain, "Recent advances in clinical oncoproteomics," *Journal of Buon*, vol. 12, supplement 1, pp. S31–S38, 2007.

[26] Y. Mao, J. Yu, J. Chen, et al., "Diagnosis of renal allograft subclinical rejection by urine protein fingerprint analysis," *Transplant Immunology*, vol. 18, no. 3, pp. 255–259, 2008.

[27] R. Martinez-Pinna, J. L. Martin-Ventura, S. Mas, L. M. Blanco-Colio, J. Tunon, and J. Egido, "Proteomics in atherosclerosis," *Current Atherosclerosis Reports*, vol. 10, no. 3, pp. 209–215, 2008.

[28] M.-A. Meuwis, M. Fillet, P. Geurts, et al., "Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling," *Biochemical Pharmacology*, vol. 73, no. 9, pp. 1422–1433, 2007.

[29] M.-A. Meuwis, M. Fillet, L. Lutteri, et al., "Proteomics for prediction and characterization of response to infliximab in Crohn's disease: a pilot study," *Clinical Biochemistry*, vol. 41, no. 12, pp. 960–967, 2008.

[30] S. J. Pitteri and S. M. Hanash, "Proteomic approaches for cancer biomarker discovery in plasma," *Expert Review of Proteomics*, vol. 4, no. 5, pp. 589–590, 2007.

[31] E. R. Sauter, S. Shan, J. E. Hewett, P. Speckman, and G. C. Du Bois, "Proteomic analysis of nipple aspirate fluid using SELDl-TOF-MS," *International Journal of Cancer*, vol. 114, no. 5, pp. 791–796, 2005.

[32] Y. Shen, J. Kim, E. F. Strittmatter, et al., "Characterization of the human blood plasma proteome," *Proteomics*, vol. 5, no. 15, pp. 4034–4045, 2005.

[33] A. H. Simonsen, J. McGuire, V. N. Podust, et al., "Identification of a novel panel of cerebrospinal fluid biomarkers for Alzheimer's disease," *Neurobiology of Aging*, vol. 29, no. 7, pp. 961–968, 2008.

[34] N. Tomosugi, K. Kitagawa, N. Takahashi, S. Sugai, and I. Ishikawat, "Diagnostic potential of tear proteomic patterns in Sjogren's syndrome," *Journal of Proteome Research*, vol. 4, no. 3, pp. 820–825, 2005.

[35] H. Zetterberg, U. Rüetschi, E. Portelius, et al., "Clinical proteomics in neurodegenerative disorders," *Acta Neurologica Scandinavica*, vol. 118, no. 1, pp. 1–11, 2008.

[36] J. Y. M. N. Engwegen, N. Mehra, J. B. A. G. Haanen, et al., "Validation of SELDI-TOF MS serum protein profiles for renal cell carcinoma in new populations," *Laboratory Investigation*, vol. 87, no. 2, pp. 161–172, 2007.

[37] D. M. Good, V. Thongboonkerd, J. Novak, et al., "Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future," *Journal of Proteome Research*, vol. 6, no. 12, pp. 4549–4555, 2007.

[38] M. Kiehntopf, R. Siegmund, and T. Deufel, "Use of SELDI-TOF mass spectrometry for identification of new biomarkers: potential and limitations," *Clinical Chemistry and Laboratory Medicine*, vol. 45, no. 11, pp. 1435–1449, 2007.

[39] D. McLerran, W. E. Grizzle, Z. Feng, et al., "Analytical validation of serum proteomic profiling for diagnosis of prostate cancer: sources of sample bias," *Clinical Chemistry*, vol. 54, no. 1, pp. 44–52, 2008.

[40] D. McLerran, W. E. Grizzle, Z. Feng, et al., "SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer," *Clinical Chemistry*, vol. 54, no. 1, pp. 53–60, 2008.

[41] A. Villar-Garea, M. Griese, and A. Imhof, "Biomarker discovery from body fluids using mass spectrometry," *Journal of Chromatography B*, vol. 849, no. 1-2, pp. 105–114, 2007.

[42] P. Maurya, P. Meleady, P. Dowling, and M. Clynes, "Proteomic approaches for serum biomarker discovery in cancer," *Anticancer Research*, vol. 27, no. 3A, pp. 1247–1255, 2007.

[43] T. L. Nickolas, J. Barasch, and P. Devarajan, "Biomarkers in acute and chronic kidney disease," *Current Opinion in Nephrology and Hypertension*, vol. 17, no. 2, pp. 127–132, 2008.

[44] L. C. Whelan, K. A. R. Power, D. T. McDowell, J. Kennedy, and W. M. Gallagher, "Applications of SELDI-MS technology in oncology," *Journal of Cellular and Molecular Medicine*, vol. 12, no. 5, pp. 1535–1547, 2008.

[45] P. Findeisen and M. Neumaier, "Mass spectrometry based proteomics profiling as diagnostic tool in oncology: current status and future perspective," *Clinical Chemistry and Laboratory Medicine*, vol. 47, no. 6, pp. 666–684, 2009.

[46] R. Apweiler, C. Aslanidis, T. Deufel, et al., "Approaching clinical proteomics: current state and future fields of application in fluid proteomics," *Clinical Chemistry and Laboratory Medicine*, vol. 47, no. 6, pp. 724–744, 2009.

[47] J. Albrethsen, R. Bogebo, J. Olsen, H. Raskov, and S. Gammeltoft, "Preanalytical and analytical variation of surface-enhanced laser desorption-ionization time-of-flight mass spectrometry of human serum," *Clinical Chemistry and Laboratory Medicine*, vol. 44, no. 10, pp. 1243–1252, 2006.

[48] R. E. Banks, A. J. Stanley, D. A. Cairns, et al., "Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry," *Clinical Chemistry*, vol. 51, no. 9, pp. 1637–1649, 2005.

[49] S. Barelli, D. Crettaz, L. Thadikkaran, O. Rubin, and J.-D. Tissot, "Plasma/serum proteomics: pre-analytical issues," *Expert Review of Proteomics*, vol. 4, no. 3, pp. 363–370, 2007.

[50] M. Dijkstra, R. J. Vonk, and R. C. Jansen, "SELDI-TOF mass spectra: a view on sources of variation," *Journal of Chromatography B*, vol. 847, no. 1, pp. 12–23, 2007.

[51] G. L. Hortin, "Can mass spectrometric protein profiling meet desired standards of clinical laboratory practice?" *Clinical Chemistry*, vol. 51, no. 1, pp. 3–5, 2005.

[52] J. N. Mcguire, J. Overgaard, and F. Pociot, "Mass spectrometry is only one piece of the puzzle in clinical proteomics," *Briefings in Functional Genomics and Proteomics*, vol. 7, no. 1, pp. 74–83, 2008.

[53] H. Mischak, R. Apweiler, R. E. Banks, et al., "Clinical proteomics: a need to define the field and to begin to set adequate standards," *Proteomics Clinical Applications*, vol. 1, no. 2, pp. 148–156, 2007.

[54] M. Papale, M. C. Pedicillo, B. J. Thatcher, et al., "Urine profiling by SELDI-TOF/MS: monitoring of the critical steps in sample collection, handling and analysis," *Journal of Chromatography B*, vol. 856, no. 1-2, pp. 205–213, 2007.

[55] T. C. W. Poon, "Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices," *Expert Review of Proteomics*, vol. 4, no. 1, pp. 51–65, 2007.

[56] A. J. Rai, C. A. Gelfand, B. C. Haywood, et al., "HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples," *Proteomics*, vol. 5, no. 13, pp. 3262–3277, 2005.

[57] A. J. Rai, P. M. Stemmer, Z. Zhang, et al., "Analysis of human proteome organization plasma proteome project (HUPO PPP) reference specimens using surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry: multi-institution correlation of spectra and identification of biomarkers," *Proteomics*, vol. 5, no. 13, pp. 3467–3474, 2005.

[58] A. J. Rai and F. Vitzthum, "Effects of preanalytical variables on peptide and protein measurements in human serum and plasma: implications for clinical proteomics," *Expert Review of Proteomics*, vol. 3, no. 4, pp. 409–426, 2006.

[59] D. Rollin, T. Whistler, and S. D. Vernon, "Laboratory methods to improve SELDI peak detection and quantitation," *Proteome Science*, vol. 5, p. 9, 2007.

[60] R. Schipper, A. Loof, J. de Groot, L. Harthoorn, E. Dransfield, and W. van Heerde, "SELDI-TOF-MS of saliva: methodology and pre-treatment effects," *Journal of Chromatography B*, vol. 847, no. 1, pp. 45–53, 2007.

[61] J. F. Timms, E. Arslan-Low, A. Gentry-Maharaj, et al., "Preanalytic influence of sample handling on SELDI-TOF serum protein profiles," *Clinical Chemistry*, vol. 53, no. 4, pp. 645–656, 2007.

[62] C. N. White, Z. Zhang, and D. W. Chan, "Quality control for SELDI analysis," *Clinical Chemistry and Laboratory Medicine*, vol. 43, no. 2, pp. 125–126, 2005.

[63] O. J. Semmes, Z. Feng, B.-L. Adam, et al., "Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility," *Clinical Chemistry*, vol. 51, no. 1, pp. 102–112, 2005.

[64] K. A. Baggerly, J. S. Morris, and K. R. Coombes, "Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments," *Bioinformatics*, vol. 20, no. 5, pp. 777–785, 2004.

[65] J. Hu, K. R. Coombes, J. S. Morris, and K. A. Baggerly, "The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales," *Briefings in Functional Genomics and Proteomics*, vol. 3, no. 4, pp. 322–331, 2005.

[66] J. Hu, K. R. Coombes, J. S. Morris, and K. A. Baggerly, "The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales," *Briefings in Functional Genomics and Proteomics*, vol. 3, no. 4, pp. 322–331, 2005.

[67] S.-Y. Hsieh, R.-K. Chen, Y.-H. Pan, and H.-L. Lee, "Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling," *Proteomics*, vol. 6, no. 10, pp. 3189–3198, 2006.

[68] C. E. L. Dammann, M. Meyer, O. Dammann, and N. von Neuhoff, "Protein detection in dried blood by surface-enhanced laser desorption/ionization-time of flight mass spectrometry (SELDI-TOF MS)," *Biology of the Neonate*, vol. 89, no. 2, pp. 126–132, 2006.

[69] F. H. Grus, V. N. Podust, K. Bruns, et al., "SELDI-TOF-MS ProteinChip array profiling of tears from patients with dry eye," *Investigative Ophthalmology and Visual Science*, vol. 46, no. 3, pp. 863–876, 2005.

[70] S. Hu, J. A. Loo, and D. T. Wong, "Human body fluid proteome analysis," *Proteomics*, vol. 6, no. 23, pp. 6326–6353, 2006.

[71] J. L. Noble, R. S. Dua, G. R. Coulton, C. M. Isacke, and G. P. H. Gui, "A comparative proteinomic analysis of nipple aspiration fluid from healthy women and women with breast cancer," *European Journal of Cancer*, vol. 43, no. 16, pp. 2315–2320, 2007.

[72] E. O'Riordan, T. N. Orlova, V. N. Podust, et al., "Characterization of urinary peptide biomarkers of acute rejection in renal allografts," *American Journal of Transplantation*, vol. 7, no. 4, pp. 930–940, 2007.

[73] M. Papale, M. C. Pedicillo, S. Di Paolo, et al., "Saliva analysis by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF/MS): from sample collection to data analysis," *Clinical Chemistry and Laboratory Medicine*, vol. 46, no. 1, pp. 89–99, 2008.

[74] R. Schipper, A. Loof, J. de Groot, L. Harthoorn, W. van Heerde, and E. Dransfield, "Salivary protein/peptide profiling with SELDI-TOF-MS," *Annals of the New York Academy of Sciences*, vol. 1098, pp. 498–503, 2007.

[75] F. J. Schweigert, B. Gericke, W. Wolfram, U. Kaisers, and J. W. Dudenhausen, "Peptide and protein profiles in serum and follicular fluid of women undergoing IVF," *Human Reproduction*, vol. 21, no. 11, pp. 2960–2968, 2006.

[76] K. S. Shores and D. R. Knapp, "Assessment approach for evaluating high abundance protein depletion methods for cerebrospinal fluid (CSF) proteomic analysis," *Journal of Proteome Research*, vol. 6, no. 9, pp. 3739–3751, 2007.

[77] M. Suzuki, G. F. Ross, K. Wiers, et al., "Identification of a urinary proteomic signature for lupus nephritis in children," *Pediatric Nephrology*, vol. 22, no. 12, pp. 2047–2057, 2007.

[78] A. Goncalves, E. Charafe-Jauffret, F. Bertucci, et al., "Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes," *Molecular and Cellular Proteomics*, vol. 7, no. 8, pp. 1420–1433, 2008.

[79] C. Jansen, K. M. Hebeda, M. Linkels, et al., "Protein profiling of B-cell lymphomas using tissue biopsies: a potential tool for small samples in pathology," *Cellular Oncology*, vol. 30, no. 1, pp. 27–38, 2008.

[80] S. Schaub, J. Wilkins, T. Weiler, K. Sangster, D. Rush, and P. Nickerson, "Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry," *Kidney International*, vol. 65, no. 1, pp. 323–332, 2004.

[81] H. Tammen, I. Schulte, R. Hess, et al., "Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display," *Proteomics*, vol. 5, no. 13, pp. 3414–3422, 2005.

[82] A. Z. Traum, M. P. Wells, M. Aivado, T. A. Libermann, M. F. Ramoni, and A. D. Schachter, "SELDI-TOF MS of quadruplicate urine and serum samples to evaluate changes related to storage conditions," *Proteomics*, vol. 6, no. 5, pp. 1676–1680, 2006.

[83] J. Yi, C. Kim, and C. A. Gelfand, "Inhibition of intrinsic proteolytic activities moderates preanalytical variability and instability of human plasma," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1768–1781, 2007.

[84] A. E. Pelzer, I. Feuerstein, C. Fuchsberger, et al., "Influence of blood sampling on protein profiling and pattern analysis using matrix-assisted laser desorption/ionisation mass spectrometry," *BJU International*, vol. 99, no. 3, pp. 658–662, 2007.

[85] J. L. Luque-Garcia and T. A. Neubert, "Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry," *Journal of Chromatography A*, vol. 1153, no. 1-2, pp. 259–276, 2007.

[86] M. Battino, M. S. Ferreiro, I. Gallardo, H. N. Newman, and P. Bullon, "The antioxidant capacity of saliva," *Journal of Clinical Periodontology*, vol. 29, no. 3, pp. 189–194, 2002.

[87] H. Voorbij, "Pouten en risiko's in een klinisch-chemisch laboratorium," *Nederlands Tijdschrift voor Klinische Chemie*, vol. 19, pp. 36–41, 1994.

[88] A. Gagnon, Q. Shi, and B. Ye, "Surface-enhanced laser desorption/ionization mass spectrometry for protein and peptide profiling of body fluids," *Methods in Molecular Biology*, vol. 441, pp. 41–56, 2008.

[89] G. MacGregor, R. D. Gray, T. N. Hilliard, et al., "Biomarkers for cystic fibrosis lung disease: application of SELDI-TOF mass spectrometry to BAL fluid," *Journal of Cystic Fibrosis*, vol. 7, no. 5, pp. 352–358, 2008.

[90] J. T.-J. Huang, L. Wang, S. Prabakaran, et al., "Independent protein-profiling studies show a decrease in apolipoprotein A1 levels in schizophrenia CSF, brain and peripheral tissues," *Molecular Psychiatry*, vol. 13, no. 12, pp. 1118–1128, 2008.

[91] N. L. Anderson and N. G. Anderson, "The human plasma proteome: history, character, and diagnostic prospects," *Molecular & Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, 2002.

[92] M. A. Gillette, D. R. Mani, and S. A. Carr, "Place of pattern in proteomic biomarker discovery," *Journal of Proteome Research*, vol. 4, no. 4, pp. 1143–1154, 2005.

[93] M. K. D. R. Dayarathna, W. S. Hancock, and M. Hincapie, "A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: application to the analysis of obesity, diabetes, and hypertension diseases," *Journal of Separation Science*, vol. 31, no. 6-7, pp. 1156–1166, 2008.

[94] X. Jiang, M. Ye, and H. Zou, "Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis," *Proteomics*, vol. 8, no. 4, pp. 686–705, 2008.

[95] P. Matt, Z. Fu, Q. Fu, and J. E. Van Eyk, "Biomarker discovery: proteome fractionation and separation in biological samples," *Physiological Genomics*, vol. 33, no. 1, pp. 12–17, 2008.

[96] M. Pernemalm, L. M. Orre, J. Lengqvist, et al., "Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery," *Journal of Proteome Research*, vol. 7, no. 7, pp. 2712–2722, 2008.

[97] P. G. Righetti, E. Boschetti, L. Lomas, and A. Citterio, "Protein equalizer technology : the quest for a "democratic proteome"," *Proteomics*, vol. 6, no. 14, pp. 3980–3992, 2006.

[98] K. S. Shores, D. G. Udugamasooriya, T. Kodadek, and D. R. Knapp, "Use of peptide analogue diversity library beads for increased depth of proteomic analysis: application to cerebrospinal fluid," *Journal of Proteome Research*, vol. 7, no. 5, pp. 1922–1931, 2008.

[99] C. Sihlbom, I. Kanmert, H. V. Bahr, and P. Davidsson, "Evaluation of the combination of bead technology with SELDI-TOF-MS and 2-D DIGE for detection of plasma proteins," *Journal of Proteome Research*, vol. 7, no. 9, pp. 4191–4198, 2008.

[100] S. W. Tam, J. Pirro, and D. Hinerfeld, "Depletion and fractionation technologies in plasma proteomic analysis," *Expert Review of Proteomics*, vol. 1, no. 4, pp. 411–420, 2004.

[101] J. R. Whiteaker, H. Zhang, J. K. Eng, et al., "Head-to-head comparison of serum fractionation techniques," *Journal of Proteome Research*, vol. 6, no. 2, pp. 828–836, 2007.

[102] F. E. Ahmed, "Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry," *Journal of Separation Science*, vol. 32, no. 5-6, pp. 771–798, 2009.

[103] N. Zolotarjova, P. Mrozinski, H. Chen, and J. Martosella, "Combination of affinity depletion of abundant proteins and reversed-phase fractionation in proteomic analysis of human plasma/serum," *Journal of Chromatography A*, vol. 1189, no. 1-2, pp. 332–338, 2008.

[104] M. Pernemalm, R. Lewensohn, and J. Lehtio, "Affinity prefractionation for MS-based plasma proteomics," *Proteomics*, vol. 9, no. 6, pp. 1420–1427, 2009.

[105] P. G. Righetti and E. Boschetti, "Sherlock Holmes and the proteome—a detective story," *FEBS Journal*, vol. 274, no. 4, pp. 897–905, 2007.

[106] L. A. Liotta and E. F. Petricoin, "Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold," *The Journal of Clinical Investigation*, vol. 116, no. 1, pp. 26–30, 2006.

[107] P. G. Righetti and E. Boschetti, "The ProteoMiner and the FortyNiners: searching for gold nuggets in the proteomic arena," *Mass Spectrometry Reviews*, vol. 27, no. 6, pp. 596–608, 2008.

[108] A. Castagna, D. Cecconi, L. Sennels, et al., "Exploring the hidden human urinary proteome via ligand library beads," *Journal of Proteome Research*, vol. 4, no. 6, pp. 1917–1930, 2005.

[109] L. Guerrier, S. Claverol, F. Fortis, et al., "Exploring the platelet proteome via combinatorial, hexapeptide ligand libraries," *Journal of Proteome Research*, vol. 6, no. 11, pp. 4290–4303, 2007.

[110] L. Sennels, M. Salek, L. Lomas, E. Boschetti, P. G. Righetti, and J. Rappsilber, "Proteomic analysis of human blood serum using peptide library beads," *Journal of Proteome Research*, vol. 6, no. 10, pp. 4055–4062, 2007.

[111] V. Thulasiraman, S. Lin, L. Gheorghiu, et al., "Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands," *Electrophoresis*, vol. 26, no. 18, pp. 3561–3571, 2005.

[112] J. Albrethsen, "Reproducibility in protein profiling by MALDI-TOF mass spectrometry," *Clinical Chemistry*, vol. 53, no. 5, pp. 852–858, 2007.

[113] L. M. Preston, K. K. Murray, and D. H. Russell, "Reproducibility and quantitation of matrix-assisted laser desorption ionization mass spectrometry: effects of nitrocellulose on peptide ion yields," *Biological Mass Spectrometry*, vol. 22, no. 9, pp. 544–550, 1993.

[114] L. Ekblad, B. Baldetorp, M. Ferno, H. Olsson, and C. Bratt, "In-source decay causes artifacts in SELDI-TOF MS spectra," *Journal of Proteome Research*, vol. 6, no. 4, pp. 1609–1614, 2007.

[115] D. A. Cairns, D. Thompson, D. N. Perkins, A. J. Stanley, P. J. Selby, and R. E. Banks, "Proteomic profiling using mass spectrometry—does normalising by total ion current potentially mask some biological differences?" *Proteomics*, vol. 8, no. 1, pp. 21–27, 2008.

[116] W. Meuleman, J. Y. M. N. Engwegen, M.-C. W. Gast, J. H. Beijnen, M. J. T. Reinders, and L. F. A. Wessels, "Comparison of normalisation methods for surface-enhanced laser desorption and ionisation (SELDI) time-of-flight (TOF) mass spectrometry data," *BMC Bioinformatics*, vol. 9, article 88, 2008.

[117] W. Wegdam, P. D. Moerland, M. R. Buist, et al., "Classification-based comparison of pre-processing methods for interpretation of mass spectrometry generated clinical datasets," *Proteome Science*, vol. 7, p. 19, 2009.

[118] N. Jeffries, "Algorithms for alignment of mass spectrometry proteomic data," *Bioinformatics*, vol. 21, no. 14, pp. 3066–3073, 2005.

[119] C. S. Tan, A. Ploner, A. Quandt, J. Lehtio, and Y. Pawitan, "Finding regions of significance in SELDI measurements for identifying protein biomarkers," *Bioinformatics*, vol. 22, no. 12, pp. 1515–1523, 2006.

[120] F. M. Smith, W. M. Gallagher, E. Fox, et al., "Combination of SELDI-TOF-MS and data mining provides early-stage

response prediction for rectal tumors undergoing multimodal neoadjuvant therapy," *Annals of Surgery*, vol. 245, no. 2, pp. 259–266, 2007.

[121] P. Geurts, M. Fillet, D. de Seny, et al., "Proteomic mass spectra classification using decision tree based ensemble methods," *Bioinformatics*, vol. 21, no. 14, pp. 3138–3145, 2005.

[122] S. Smit, H. C. J. Hoefsloot, and A. K. Smilde, "Statistical data processing in clinical proteomics," *Journal of Chromatography B*, vol. 866, no. 1-2, pp. 77–88, 2008.

[123] E. T. Fung, T.-T. Yip, L. Lomas, et al., "Classification of cancer types by measuring variants of host response proteins using SELDI serum assays," *International Journal of Cancer*, vol. 115, no. 5, pp. 783–789, 2005.

[124] J. Peng, A. J. Stanley, D. Cairns, P. J. Selby, and R. E. Banks, "Using the protein chip interface with quadrupole time-of-flight mass spectrometry to directly identify peaks in SELDI profiles—initial evaluation using low molecular weight serum peaks," *Proteomics*, vol. 9, no. 2, pp. 492–498, 2009.

[125] R. N. H. Pugh, I. M. Murray Lyon, J. L. Dawson, M. C. Pietroni, and R. Williams, "Transection of the oesophagus for bleeding oesophageal varices," *British Journal of Surgery*, vol. 60, no. 8, pp. 646–649, 1973.

[126] L. Brugler, A. Stankovic, L. Bernstein, F. Scott, and J. O'Sullivan-Maillet, "The role of visceral protein markers in protein calorie malnutrition," *Clinical Chemistry and Laboratory Medicine*, vol. 40, no. 12, pp. 1360–1369, 2002.

[127] P. B. Sehgal, "Interleukin-6: molecular pathophysiology," *Journal of Investigative Dermatology*, vol. 94, no. 6, supplement, pp. 2S–6S, 1990.

[128] E. P. Diamandis, "Point: proteomic patterns in biological fluids: do they represent the future of cancer diagnostics?" *Clinical Chemistry*, vol. 49, no. 8, pp. 1272–1275, 2003.

[129] N. Escher, M. Kaatz, C. Melle, et al., "Posttranslational modifications of transthyretin are serum markers in patients with mycosis fungoides," *Neoplasia*, vol. 9, no. 3, pp. 254–259, 2007.

[130] M. Ueda, Y. Misumi, M. Mizuguchi, et al., "SELDI-TOF mass spectrometry evaluation of variant transthyretins for diagnosis and pathogenesis of familial amyloidotic polyneuropathy," *Clinical Chemistry*, vol. 55, no. 6, pp. 1223–1227, 2009.

[131] L. Chen, S. Fatima, J. Peng, and X. Leng, "SELDI protein chip technology for the detection of serum biomarkers for liver disease," *Protein and Peptide Letters*, vol. 16, no. 5, pp. 467–472, 2009.

[132] B. Matharoo-Ball, G. Ball, and R. Rees, "Clinical proteomics: discovery of cancer biomarkers using mass spectrometry and bioinformatics approaches—a prostate cancer perspective," *Vaccine*, vol. 25, supplement 2, pp. B110–B121, 2007.

[133] A. Skytt, E. Thysell, P. Stattin, et al., "SELDI-TOF MS versus prostate specific antigen analysis of prospective plasma samples in a nested case-control study of prostate cancer," *International Journal of Cancer*, vol. 121, no. 3, pp. 615–620, 2007.

[134] G. Xu, C. Q. Xiang, Y. Lu, et al., "Application of SELDI-TOF-MS to identify serum biomarkers for renal cell carcinoma," *Cancer Letters*, vol. 282, no. 2, pp. 205–213, 2009.

[135] B. J. McMorran, S. A. O. Patat, J. B. Carlin, et al., "Novel neutrophil-derived proteins in bronchoalveolar lavage fluid indicate an exaggerated inflammatory response in pediatric cystic fibrosis patients," *Clinical Chemistry*, vol. 53, no. 10, pp. 1782–1791, 2007.

[136] C. Rocken, R. Ketterlinus, and M. P. A. Ebert, "Application of proteome analysis to the assessment of prognosis and

response prediction in clinical oncology," *Current Cancer Drug Targets*, vol. 8, no. 2, pp. 141–145, 2008.

[137] C. Trocme, H. Marotte, A. Baillet, et al., "Apolipoprotein A-I and platelet factor 4 are biomarkers for infliximab response in rheumatoid arthritis," *Annals of the Rheumatic Diseases*, vol. 68, no. 8, pp. 1328–1333, 2009.

Review Article

Insights into the Biology of IRES Elements through Riboproteomic Approaches

Almudena Pacheco[1, 2] and Encarnacion Martinez-Salas[1]

[1] *Centro de Biologia Molecular Severo Ochoa, CSIC-UAM, Nicolas Cabrera 1, Cantoblanco, 28049 Madrid, Spain*
[2] *Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA*

Correspondence should be addressed to Encarnacion Martinez-Salas, emartinez@cbm.uam.es

Translation initiation is a highly regulated process that exerts a strong influence on the posttranscriptional control of gene expression. Two alternative mechanisms govern translation initiation in eukaryotic mRNAs, the cap-dependent initiation mechanism operating in most mRNAs, and the internal ribosome entry site (IRES)-dependent mechanism, first discovered in picornaviruses. IRES elements are highly structured RNA sequences that, in most instances, require specific proteins for recruitment of the translation machinery. Some of these proteins are eukaryotic initiation factors. In addition, RNA-binding proteins (RBPs) play a key role in internal initiation control. RBPs are pivotal regulators of gene expression in response to numerous stresses, including virus infection. This review discusses recent advances on riboproteomic approaches to identify IRES transacting factors (ITAFs) and the relationship between RNA-protein interaction and IRES activity, highlighting the most relevant features on picornavirus and hepatitis C virus IRESs.

1. Translational Control of Gene Expression

The composition of the cellular proteome at any given time is tightly regulated. This is achieved by fine-tuning of all the processes governing gene expression, including transcription, splicing, mRNA transport, RNA stability, translation, protein stability and posttranslational modification. All the steps within this cascade of events are subjected to their own specific regulation, and contribute to generate a different composition of the proteome by modifying not only the levels but also the identity of the proteins present in the cell under specific conditions. The components that participate in these regulatory events are often engaged in the formation of macromolecular complexes. Protein-protein as well as RNA-protein interactions allow a compartmentalization of the factors needed to control gene expression. Translational control is a key determinant of the cellular proteome. Translation initiation can modify the proteome by altering the efficiency of translation as well as by enabling the synthesis of different forms of a protein from specific genes.

The process of RNA translation includes a series of sequential steps, known as initiation, elongation, termination and ribosome recycling. Most of translational control is exerted at the initiation step, assisted by specific proteins designated translation initiation factors (eIFs). Translation initiation of most eukaryotic mRNAs commences with 5′ end-dependent recruitment of the 43S complex (that is composed of a 40S subunit bound to eIF2-GTP/Met-tRNAi, eIF1A, eIF1 and eIF3) by eIF4F recognition of the m^7GpppN (cap) at the 5′ end of the mRNA. In turn, the eIF4F complex comprises eIF4E that physically binds to the cap, eIF4A that unwinds secondary structures in the 5′ untranslated region (5′UTR) and eIF4G, a scaffold protein that interacts with eIF4E, eIF4A and eIF3. Aided by the helicase activity of eIF4A and its cofactor eIF4B, the 43S pre-initiation complex scans in 5′ to 3′ direction until an appropriate initiation codon is encountered. Auxiliary factors, eIF1, eIF2 and eIF5, help to identify the correct AUG start codon, resulting in the formation of the 48S complex. eIF5 induces hydrolysis of eIF2-bound GTP, which is recycled to the active form by

eIF2B (guanine nucleotide exchange factor). The poly(A) tail present in the 3′UTR of most mRNAs synergistically stimulates the efficiency of translation through recruitment of poly(A)-binding protein (PABP), enabling its interaction with eIF4F located at the mRNA 5′ end. Finally, eIF5B mediates joining of the 60S and 40S subunits, generating the 80S complex competent for protein synthesis [1].

Translation initiation, particularly in viral mRNAs, can occur by an alternative mechanism driven by internal ribosome entry site (IRES) elements, discovered near 20 years ago in two picornaviruses, encephalomyocarditis virus (EMCV) and poliovirus (PV) [2, 3]. These elements are *cis*-acting sequences that form secondary and tertiary RNA structures and recruit the translation machinery to an internal position in the mRNA, bypassing a large number of stable structural elements in the viral 5′UTR [4]. Hence, picornavirus IRES-driven initiation is 5′ end-independent and does not require eIF4E to recruit the 40S subunit, in contrast to the cap-dependent initiation mechanism.

The subsequent discovery of an IRES element in hepatitis C virus (HCV) RNA that was able to recruit 40S ribosomal subunits in the absence of eIF4G represented a major breakthrough in the translation field [5, 6]. This finding opened the question of how IRESs differing in primary sequence, RNA structure, and factor requirements, perform the same function. Over the last two decades, the process of internal initiation has been found to be more general than originally thought, that is, it operates in other RNA and DNA viruses as well as in cellular mRNAs [7]. In all cases, IRESs direct translation of a subset of proteins when cap-dependent translation is severely compromised [4].

The excellent performance of IRESs, together with the fact that they are active in genetically engineered constructs, has been exploited to study how these diverse RNA elements perform the same function. With the exception of one or more polypyrimidine tracts, no primary sequence conservation neither overall structural similarity is detected between picornavirus and HCV IRESs, strongly suggesting that different strategies may be used to recruit the ribosomal subunits. These strategies could be under the control of a distinct group of proteins specifically interacting with each of these regulatory elements. In this review we discuss recent advances in the identification of IRES auxiliary factors. Understanding the role played by these IRES trans-acting factors (ITAFs) may help to unravel the strategies employed by mRNAs to capture the translation machinery internally.

2. Internal Initiation of Translation in Picornavirus RNAs

Translation initiation of all picornavirus RNAs is dependent on the IRES located in the long 5′UTR (Figure 1(a)). Various structural elements in the 5′UTR region, which differ among picornavirus genera, control the viral replication cycle in concerted action with the 3′UTR [8–11]. The IRES region spans about 450 nucleotides immediately upstream of the functional translation start codon of the polyprotein [2, 3, 12, 13]. Although less than 50% of primary sequence is

conserved between different members of the picornavirus family, the similarity of their secondary structure allows their classification into four types, I to IV. Type I includes enterovirus (EV, PV, HRV), type II, cardiovirus (EMCV) and aphthovirus (foot-and-mouth disease virus, FMDV), type III, is used for hepatitis A virus (HAV), and the HCV-like IRES conforms group IV. The acquisition of a proper structural organization is a key determinant of internal translation initiation driven by picornavirus IRES [14, 15], and all viral IRES in general [16, 17]. This feature is well illustrated by mutational and structural studies conducted on the central domain (termed 3 or I) of type II IRES (Figure 1(a)). This region is organized as a cruciform structure with phylogenetically conserved structural motifs that are essential for IRES activity [18, 19] and determine the tertiary structure of this region [20–23].

Picornavirus IRES-driven translation initiation depends on the recognition of the IRES by specific cellular proteins [4]. IRESs belonging to types I and II require eIF4G, eIF4A, eIF2, eIF3 and ATP, but no eIF4E, eIF1 or eIF1A to assemble 48S complexes in a reconstitution assay with purified components [13, 24, 25]. Specific structural motifs in the stem-loops J-K-L (or 4-5) provide the preferential binding site for eIF4G, eIF4B and eIF3 (Figure 1(a)) [26–28]. However, interaction of these eIFs is necessary but not sufficient to promote IRES activity [8], demonstrating the essential function of domains 2 and 3 in IRES function. The contribution of domains 2 and 3 to IRES activity may consist in the acquisition of a proper RNA structural organization, assisted by auxiliary proteins. Along this line, while the requirement for eIFs is well established, the ITAFs involved in picornavirus IRES activity are still under study. Here we review the RNA-binding proteins (RBPs) that can form ribonucleoprotein (RNP) complexes with IRESs modulating their efficiency of translation.

2.1. Picornavirus ITAFs. Picornavirus RNAs differ on their ability to operate in the cell-free rabbit reticulocyte lysates (RRL). Early studies conducted on the PV RNA, which was inactive in RRL, evidenced that its translation efficiency was greatly enhanced upon supplementation of the lysates with HeLa cell extracts [29]. This difference was related to the requirement of factors that were missing or present in limiting amounts in RRL. Hence, addition of HeLa cells soluble extract resulted in reconstitution of PV RNA translation [30]. In due course, these observations led to the discovery of auxiliary proteins behaving as IRES trans-acting factors. Most ITAFs described so far are well characterized RBPs that contain RNA-binding motifs organized in a modular structure [31, 32], as it also occur in proteins involved in RNA processing and transport. However, modulation of IRES activity by ITAFs is not well understood, and at least in part it is a controversial issue because, depending on the IRES element, some proteins show a more stringent requirement than others do [33, 34].

Over the last decade, the interaction of RBPs with picornavirus IRES (Table 1) has been analyzed taking advantage of riboproteomic affinity methods. Of interest, and confirming the validity of this methodology, proteins

TABLE 1: RNA-binding proteins interacting with viral IRES.

ITAFS	Effect on IRES activity	IRES targets	Functions in RNA metabolism
PTB	Stimulation	FMDV, EMCV, TMEV, PV, HRV, HAV, HCV	plicing, polyadenylation, RNA stability, localization
DRBP76:NF45	Repressor	HRV, HCV	Transcription, RNA transport, stability, viral replication
Ebp1/ PA2G4/ ITAF$_{45}$	Stimulation	FMDV, EMCV*	Transcription regulator
Unr	Stimulation	PV, HRV, HCV	Translation control
HSC70		FMDV, HCV	Viral replication
SRp20	Stimulation	PV	Splicing
PCBP2	Stimulation	PV, HRV, HAV, CVB3, FMDV*, EMCV*, HCV	RNA stability, translation control
Gemin5	Downregulation	FMDV, HCV	RNA-binding factor of SMN complex
hnRNP U		FMDV	RNA processing, stability, transcription
hnRNP K		FMDV	Transcription, RNA stability, translation control
DAZ1		FMDV	Translation control
G3BP		FMDV	Stress granules assembly
Gpiap1		FMDV, HCV	Transcription regulator, viral replication
Nucleolin	Stimulation	PV, HRV, FMDV, HCV	Ribosomal RNA maturation, transport
FBP2	Repressor	EV71	RNA stability
eIF2C/Ago		FMDV	Silencing
DHX9		HRV, FMDV, HCV	RNA helicase
DDX1		FMDV	RNA helicase
RACK1		HCV	Ribosomal subunits joining
IGF2BP1	Stimulation	HCV	RNA localization, stability, translation control
La	Stimulation	PV, EMCV, HAV#	Transcription, translation control
NSAP1/ hnRNP Q	Stimulation	HCV	RNA stability, translation control, SGs component
hnRNP L	Stimulation	HCV	RNA stability, translation control
hnRNP D	Stimulation	HCV	RNA stability, translation control
hnRNPA/B		EV71, HCV	RNA processing, translation control
GAPDH	Repressor	HAV	RNA transport, translation control
YB-BP1		HCV	Transcription, RNA stability, translation control

*No effect, #suppression.
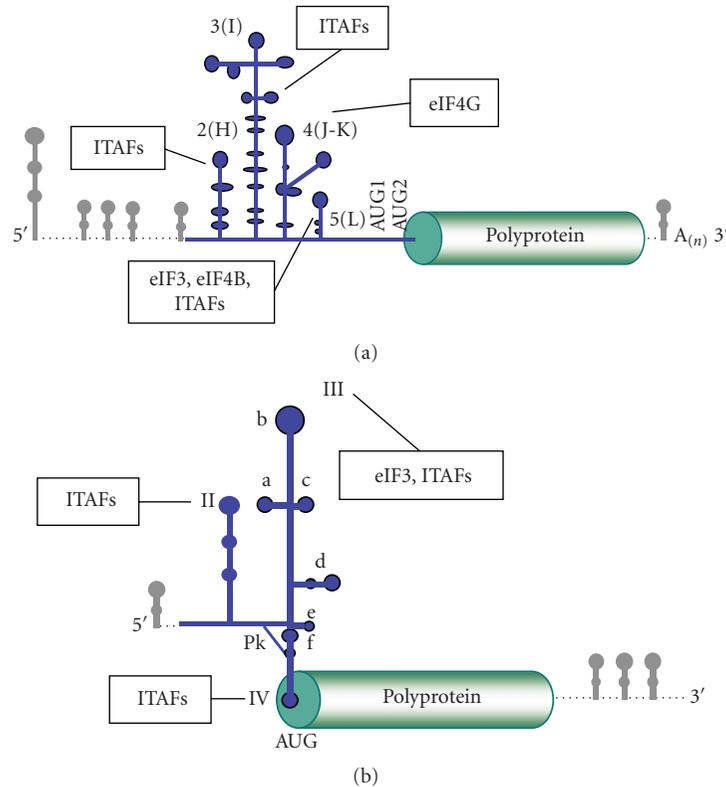
FIGURE 1: (a) Schematic representation of the picornavirus genome, using as example the foot-and-mouth disease viral RNA. (b) Schematic representation of the hepatitis C virus genome. The IRESs are depicted in blue, with indication of the domains referred to in the text. The position of the functional initiator AUGs, as well as the preferential binding sites of eIFs and ITAFs are indicated. Stable stem-loops located at the 5′ and 3′ end of the viral RNA are depicted in grey.

previously know to interact with IRESs by other methods were retrieved with the corresponding target RNA following affinity purification. This is the case of eIF4B and eIF3, which were specifically identified bound to FMDV or HCV IRES by mass spectrometry following RNA affinity purification [35–38].

Soon after the discovery of picornavirus IRESs, a direct interaction of the polypyrimidine tract-binding protein (PTB) with EMCV and FMDV IRESs was shown by UV-crosslinking [39–41], and later by RNA foot-printing [42] and hydroxyl radical probing [43]. PTB is a multifunctional RBP with four RNA recognition motifs (RRM) [31] that belongs to the heterogeneous nuclear ribonucleoprotein (hnRNP) family. The RRM domains of PTB recognize U-rich loops on short stems and in general, U/C-rich sequences [44]. This protein was originally identified in splicing-associated RNP complexes. However, it also performs critical roles in cellular processes pertaining RNA metabolism, including polyadenylation, mRNA stability and translation initiation. Regarding its function as ITAF, PTB stimulates picornavirus and retrovirus IRESs [45], but it represses translation initiation driven by the BiP IRES [46].

As a consequence of their role as regulator of IRES activity, ITAFs can mediate cell type specificity, and hence, determine viral spread. This property was brought about by the effects of the neural form of PTB (nPTB), that determines

the neurovirulence of Theiler's murine encephalitis virus (TMEV) [47], and by the double-stranded RNA-binding protein 76 (DRBP76, also termed NF90/NFAR-1), that forms a heterodimer with NF45 (nuclear factor of activated T cells). The DRBP76:NF45 heterodimer binds to the HRV2 IRES and differs in subcellular distribution in neuronal and non-neuronal malignant cells, arresting HRV translation in neuronal cells but not in glioma [48, 49].

Picornavirus IRESs often contain more than one polypyrimidine tract located in distant domains at each end of the IRES region [8]. Recent studies have shown that a single PTB molecule binds in a unique orientation to the EMCV IRES, with RRM1-2 contacting the 3′ end, and RRM3 contacting the 5′ end of the IRES, thereby constraining the IRES structure in a unique orientation [43]. However, studies carried out on the FMDV IRES raised the conclusion that RRM3-4 of PTB were bound in an oriented way to domain 2 and the IRES 3′ region, respectively [50]. Although the RRMs involved in the IRES-PTB interactions are significantly different between these two studies, both are consistent with a role of PTB in stabilizing the IRES structure, thereby acting as a chaperone.

ITAFs, as it is the case of PTB, do not act alone but in combination with various factors presumably contributing to explain the opposite effects on IRES activity [34]. Hence, RBPs interacting with different targets may result in different

effects depending on the target and the other partners of the complex. For instance, two IRESs such as EMCV and FMDV with apparent similar secondary structure but different primary sequence, exhibit different requirements in terms of functional RNA-protein association. One example is Ebp1 protein (erbB-3-binding protein 1), also known as proliferation-associated factor PA2G4 and ITAF$_{45}$, identified interacting with domain 3 of FMDV IRES in proteomic analysis [36]. Ebp1 cooperates with PTB to stimulate FMDV IRES activity in reconstitution studies [13, 51], but its depletion does not produce any effect on EMCV IRES activity [52]. This protein is expressed in proliferating cells during the S phase but not during cell cycle arrest consistent with the fact that FMDV IRES is active in proliferating tissues.

Another example of a factor that mediates IRES activity is Unr (upstream of N-ras), a cold-shock RBP that interacts with PABP1 and stimulates HRV and PV translation through its interaction with two distinct IRES domains [53, 54]. Mass spectrometry analysis identified Unr associated to the HCV IRES RNP complexes [55]. In support of the specific role of Unr in internal initiation, IRES activity of c-myc, Apaf-1, unr and PITSLRE cellular mRNAs is differentially regulated depending on the Unr-partners, hnRNP K/poly r(C)-binding protein PCBP1-2, nPTB, or hnRNP C1-2, respectively [56–58]. Other example of RBP identified with IRESs is the constitutive heat shock protein HSC70 [36], although the possibility of an indirect binding can not be discarded. HSC70 forms part of RNP complexes that interact with AU-rich elements in the 3′UTR of specific mRNAs, enhancing their stability [59].

In addition to PTB, several proteins implicated in RNA splicing can function as ITAFs, suggesting the existence of a network of interactions between different gene expression processes. An illustrative example is the splicing factor SRp20 that up-regulates PV IRES-mediated translation via its interaction with PCBP2 [60]. Another example of an ITAF involved in a different gene expression process is Gemin5 that binds directly to FMDV and HCV IRES, acting as a downregulator of translation [61]. Not surprisingly, Gemin5 is associated with other factors in RNP complexes that perform rather different roles during gene expression control. Gemin5 is the RNA-binding factor of the survival of motor neurons (SMN) complex [62], which assembles Sm proteins on snRNAs playing a critical role in the biogenesis of key components of the mRNA splicing machinery, the small nuclear ribonucleoproteins (snRNPs) [63]. Gemin5 is a nuclear protein, but it is predominantly located in the cell cytoplasm and it also appears to be present in P bodies [64].

Together, the conclusions derived from the study of multifunctional proteins such as PTB, PCBP2, Gemin5 and other ITAFs, suggest a novel mechanism for the coordinated regulation of translation initiation of a subset of mRNAs bound by shuttling proteins such as hnRNPs or splicing factors. In support of this, the splicing factor SF2/ASF mediates post-splicing activities promoting translation initiation by suppressing the activity of 4E-BP [65] and modulating the internal initiation of cellular mRNAs. In fact, it has been suggested that some RBPs might exert its

function in translation control by binding to the IRES of specific cellular mRNAs during splicing complex assembly before nuclear export [34]. This could be an additional layer of posttranscriptional regulation for proteins whose functions are important when cap-dependent translation is compromised.

hnRNPs are a family of proteins (named from hnRNP A1 to hnRNP U) with RNA-binding and protein-protein binding motifs [31, 66]. They have a nuclear localization associated with nascent RNA polymerase II transcripts and shuttle with the RNA to the cytoplasm [67]. The RGG RNA-binding motif that mediates the interaction with RNA as well as with other hnRNPs was first described in hnRNP U [66], one of the factors identified by mass spectrometry interacting with the FMDV IRES [36]. Both hnRNP U and Gemin5 form part of a complex with eIF4E, that may explain the down-regulatory role of Gemin5 in cap-dependent translation [61] by virtue of eIF4E sequestration or its localization to P bodies [64].

Several members of the hnRNP family, hnRNP K, PCBP1 (hnRNP E1) and PCBP2 (hnRNP E2), have been identified associated with various IRESs (Table 1). These proteins have in common the KH RNA binding domain first described in hnRNP K and, subsequently, in PCBP1, 2, 3 and 4 [68]. hnRNP K is the most abundant member of the family of proteins that recognize poly(rC) regions, and regulates transcription, RNA turnover and translation [69, 70]. Proteins hnRNP K, PCBP1 and PCBP2, together with DAZ-1, have been identified associated to domain 3 of the FMDV IRES [36]. DAZ1 is a 3′UTR-binding protein that has been found bound to polysomes and stimulates translation initiation of polyadenylated mRNA [71].

PCBP2 interacts with a C-rich loop in stem-loop IV of PV, CVB3 and HRV IRESs and specifically stimulates their activity [72–75]. In contrast, the activity of EMCV and FMDV IRESs that also interact with PCBP2 was not modified by the addition of recombinant PCBP2 protein to depleted-RRL lysates [76], in agreement with the lack of effect of nucleotide substitutions in the C-rich loop of FMDV IRES [77]. PCBP2 performs a dual role in translation initiation and RNA replication of the poliovirus genome [78] through its interaction with different targets in the viral 5′UTR. Furthermore, consistent with its relevance in IRES function, PV IRES competes out with the HAV IRES for PCBP2 binding [79].

The balance between translation initiation and silencing depends on the cellular response to stress. Indeed, many viruses regulate the assembly or disassembly of stress granules (SGs) modifying translation of host and virus-encoded mRNAs. Consistent with this observation, some RBPs have been localized in SGs, as PABP1 [80], or cytoplasmic processing bodies (PBs), as PCBP2 [81]. Thus, in response to stress signals including viral infection, these multifunctional proteins may perform distinct roles depending on their localization. The signaling factor Ras-GTPase-activating protein (G3BP) that was identified interacting with the FMDV IRES [36], belongs to a new family of RBPs that link tyrosin kinase-mediated signals with RNA metabolism [82]. G3BP-1 localize in cytoplasmic RNA granules [83] contributing

to its assembly [84]. These cytoplasmic aggregates contain stalled translation preinitiation complexes thought to serve as sites of mRNA storage during the cell stress response. G3BP has been found associated to the 3′UTR of HCV RNA, and its depletion induced a reduction of both HCV RNA and proteins, supporting the idea that it might be a component of HCV replicating complexes [85]. Interestingly, G3BP interacts with the transcriptional regulator GP1-anchored membrane protein (Gpiap1) also identified as an IRES-binding protein.

Many ITAFs are predominantly nuclear proteins that localize to the cytoplasm in picornavirus-infected cells [86]. Nucleolin is a protein involved in rDNA transcription, rRNA maturation, ribosome assembly and nucleo-cytoplasmic transport [87], and is translocated into the cytoplasm following infection of cells with poliovirus [88]. Nucleolin interacts with HRV, FMDV and PV IRES and stimulates PV IRES-mediated translation in transfected cells overexpressing the full-length protein [89]. During enterovirus EV71 infection, the nuclear protein FBP2 (far upstream element (FUSE) binding protein 2) was enriched in the cytoplasm where viral replication occurs, whereas in mock-infected cells FBP2 was localized in the nucleus. FBP2 is a KH protein that negatively regulates EV71 IRES activity [90] presumably through its capacity to compete with PTB binding.

Together with hnRNPs, a group of proteins that are involved in gene silencing, transport, and stabilization (eIF2C, RNA helicases) have been identified in riboproteomic approaches bound to different IRESs (Table 1). The recurrent identification of a subset of factors with different RNA targets [91, 92] points to the existence of a network of RNPs with the potential for multiple levels of regulation. Moreover, the modular structure of RBPs that is at the basis of their capacity to recognize a large number of targets raises the possibility that binding to any particular RNA could facilitate different sorts of regulation depending on the other protein partners and the cellular environment.

3. Internal Initiation of Translation in HCV RNA

Initiation of protein synthesis in the positive-strand RNA genome of HCV is also driven by an IRES [93]. The IRES region spans 340 nucleotides and differs from picornavirus IRESs in RNA structural organization and factor requirement [94]. The HCV IRES is organized in three conserved structural domains, termed II, III and IV (Figure 1(b)) that adopt a tertiary fold whose integrity is required for efficient protein synthesis [95]. Domain II, which consists of a hairpin with basal and apical loops, is essential for HCV IRES activity. Its deletion reduces translation initiation by blocking 80S formation [96]. This domain promotes eIF5-induced GTP hydrolysis during 80S ribosome assembly and eIF2/GDP release from the initiation complex [97].

Domain III consists of six stem-loops (designated IIIabcdef). The basal portion of domain III forms the core of the high-affinity interaction with the 40S subunit including

a small stem-loop (IIIe) and a pseudoknot (IIIf) [16]. In the absence of eIFs, the HCV IRES can establish a high-affinity interaction with ribosomal 40S subunits through the binding surface between subdomains IIIabc, IIIef and IIId [98]. Despite the capacity to form binary complexes, localization of the met-tRNA$_i$ on the surface of the 40S subunit by eIF2 is essential for translation, and eIF3 significantly enhances formation of the 48S initiation complex interacting with the junction of subdomains IIIabc [5, 99]. Interaction of eIF3 subunits with HCV IRES has been analyzed by cryo-electron microscopy of the binary IRES-eIF3 complex [100] and by mass spectrometry of IRES-bound protein complexes [36, 37]. However, under conditions of inactivation of eIF2 by phosphorylation, the HCV IRES can form a preinitiation complex in the presence of eIF3 and eIF5B [101], which is reminiscent of the bacterial-like initiation mode. The ribosomal proteins that participate in IRES-40S interaction have been identified by different approaches as well, such as cross-linking studies [102–104] and mass spectrometry [105].

Besides ribosomal proteins and eIF3 subunits, the non-canonical factors RACK1 and nucleolin were identified in native and IRES-40S ribosomal complexes [37]. RACK1 functions as the receptor for activated protein kinase C, and regulates translation initiation by recruiting protein kinase C to the 40S subunit. It forms a stable complex with the 40S subunit, exposing the WD-repeats as a platform for interactions with other proteins to the ribosome [106].

Another non-canonical host factor, the insulin-like growth factor II mRNA-binding protein 1 (IGF2BP1) has been reported to associate with both the IRES and the 3′UTR of HCV, and remarkably, to coimmunoprecipitate with eIF3 and the 40S subunit [38]. This result suggests that this factor may enhance HCV IRES-dependent translation by recruiting the ribosomal subunits to a pseudo-circularized RNA. In agreement with this possibility, a subset of the identified proteins, NF90/NF45, also interact with the ends of the viral RNA contributing to enhance viral RNA replication [107]. Long-range 3′–5′ interactions have been reported in the HCV viral RNA [108]. Moreover, in support of the role of the 3′–5′ interactions in the control of gene expression in positive-strand RNA viruses, stimulation of IRES activity by the homologous 3′UTR has been shown in FMDV and HRV [10, 11], presumably mediated by functional bridges that bring together the RNA ends by long-range RNA-RNA interactions [109].

Despite some controversy regarding the effect of PTB, most of the identified ITAFs regulate HCV IRES activity in a positive manner [110, 111]. La and NSAP1/hnRNP Q proteins stimulate HCV IRES-dependent translation [112]. The La autoantigen is involved in RNA polymerase III transcription. La binds to PV, EMCV and HCV IRES stimulating translation [113–115], but it suppresses HAV IRES activity [116]. NSAP1 protein has a dual function in HCV life cycle participating in RNA replication [117] and enhancing IRES-dependent translation through its binding to A-rich sequences in the core coding region [112].

Similar to NSAP1, hnRNP L interacts with the 3′ border of the HCV IRES in the core-coding sequence [118] and it is required for IRES-mediated translation [119]. This protein is necessary for efficient translation of the Cat-1 arginine/lysine transporter mRNA during amino acid starvation [120]. Other HCV-interacting protein is hnRNP D that binds to the stem-loop II and promotes translation [121]. Proteins of this family, hnRNP A/B 38, have been identified interacting with dIII of HCV IRES [36]. hnRNP A1 binds to the 5′UTR of EV71 and Sindbis RNA and is required for viral RNA replication [122]. This protein also mediates internal initiation of FGF-2 and Apaf-1 mRNAs [123, 124].

In addition to direct RNA-protein interactions, protein-protein association between RBPs, such as hnRNP U or hnRNP A/B [125, 126] during mRNA transport can explain the identification of proteins belonging to the cytoskeleton machinery with FMDV and HCV IRES [36, 38, 55]. Protein-protein interactions may also explain the identification of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with HAV IRES [127]. This protein competes with PTB for binding to stem-loop IIIa, suppressing the ability of the HAV 5′UTR to direct cap-independent translation [128]. GAPDH forms a macromolecular complex that binds to U-rich sequences in the 3′UTR of a selective group of cellular mRNAs controlling their translation [129]. However, as already mentioned for some factors, indirect interactions may be behind the identification of very abundant RBPs, such as YB-1, in riboproteomic studies. Thus, the functional involvement of each factor as well as whether the binding is direct or mediated by another partner in the RNP complex, needs to be verified individually.

A few proteins identified by mass spectrometry with a discrete domain of the HCV IRES have been also identified in similar approaches interacting with the entire HCV IRES, giving additional information about the binding site of the protein. This could be the case of RNA helicase DEAH-box polypeptide 9 (DHX9) or DEAD-box polypeptide 1 (DDX1). The DDX/DHX family of proteins play important roles in nucleic acid metabolism, including pre-mRNA processing, ribosome biogenesis, RNA turnover, RNA export, translation, and association/dissociation of large RNP complexes [130]. DHX9 recognizes a complex structure at the 5′-UTR of retrovirus mRNA precursors, facilitating its association to polyribosomes [131]. RNA chromatography assays showed that it is associated to HRV IRES [48]. DDX1, a dual interactor of hnRNP K and poly(A)-mRNA [132], has been also identified bound to the 3′UTR of HCV suggesting a role for this protein in viral RNA replication [85].

4. Concluding Remarks

In general, ITAFs are RNA-binding proteins that shuttle between the nucleus and the cytoplasm. Thus, a network of RNA-protein and protein-protein interactions may assist to recruit the IRES to the ribosome and possibly, to other cytoplasmic structures. RBPs are key cellular components that control the temporal, spatial and functional dynamics of RNA within the competitive cell environment. Indeed, changes in the expression of RBPs have profound implications for cellular physiology, affecting RNA processes from pre-mRNA splicing to protein translation [32]. Thereby, the composition of RNP complexes bound to the RNA in a particular situation will determine the fate of the RNA (e.g., stability, translatability, compartmentalization). In other words, binding of proteins, even those considered to be promiscuous, to a given RNA could mediate specific regulation. In agreement with this, recent mass spectrometry identification of the proteins associated with Argonaute proteins, the protein complex responsible for gene-silencing pathways guided by small RNAs, revealed a common group of helicases, hnRNPs and other RBPs which are shared with RNP complexes involved in other cellular processes such as mRNA transport, stabilization and translation [92].

The observation that proteins with the potential for multiple levels of regulation can recognize various RNA targets raises the possibility that protein-binding to specific RNAs could facilitate different sorts of regulation depending on the other partners and the cellular environment. Thus, elucidating the function of ITAFs will require a deep understanding of their RNA targets, their protein partners, and their potential modifications. Concerning the first issue, the recent advances in cross-linking immunoprecipitation and high-throughput sequencing appears to be a promising technique to help in this task. Implementation of new proteomic approaches will continue to help in the second and third tasks. Finally, regarding the modification of RBPs in infected cells, understanding the effect of proteolytic cleavage of factors such as PCBP2, PTB, PABP or G3BP [78, 133, 134] will need to be extended to newly identified ITAFs. All together, this will help to understand the integrated action of ITAFs on mRNA targets.

The RBPs modulating picornavirus and HCV IRES activity offer promising targets to combat these infectious pathogens. Indeed, IRESs are ideal candidates to interfere with virus replication through direct IRES-targeting or through ITAF-targeting. In the first case, antiviral agents based on RNA molecules aimed to disrupt the IRES structure have been partially successful [135–138]. In the second case, knowledge of the structural organization of ITAFs provided the basis to design antiviral therapy, as shown by a synthetic peptide derived from the RRM2 of La which acts as a dominant negative inhibitor of HCV RNA translation [113]. In the coming years, elucidation of the structural determinant of peptides derived from different ITAFs, interfering with the capacity of these proteins to generate protein-protein and RNA-protein networks, will provide the basis for developing small peptidomimetic structures as potent antiviral therapeutics.

Acknowledgments

References

[1] T. V. Pestova, J. R. Lorsch, and C. U. T. Hellen, "The mechanism of translation initiation in eukaryotes," in *Translation Control in Biology and Medicine*, M. B. Mathews, N. Sonenberg, and J. W. B. Hershey, Eds., pp. 87–128, Cold Spring Harbor laboratory, Cold Spring Harbor, NY, USA, 2007.

[2] S. K. Jang, H.-G. Krausslich, M. J. H. Nicklin, G. M. Duke, A. C. Palmenberg, and E. Wimmer, "A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation," *Journal of Virology*, vol. 62, no. 8, pp. 2636–2643, 1988.

[3] J. Pelletier and N. Sonenberg, "Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA," *Nature*, vol. 334, no. 6180, pp. 320–325, 1988.

[4] E. Martínez-Salas, A. Pacheco, P. Serrano, and N. Fernandez, "New insights into internal ribosome entry site elements relevant for viral gene expression," *Journal of General Virology*, vol. 89, no. 3, pp. 611–626, 2008.

[5] T. V. Pestova, I. N. Shatsky, S. P. Fletcher, R. J. Jackson, and C. U. T. Hellen, "A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initation of hepatitis C and classical swine fever virus RNAs," *Genes and Development*, vol. 12, no. 1, pp. 67–83, 1998.

[6] C. M. T. Spahn, J. S. Kieft, R. A. Grassucci, et al., "Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit," *Science*, vol. 291, no. 5510, pp. 1959–1962, 2001.

[7] S. D. Baird, M. Turcotte, R. G. Korneluk, and M. Holcik, "Searching for IRES," *RNA*, vol. 12, no. 10, pp. 1755–1785, 2006.

[8] O. Fernández-Miragall, S. López de Quinto, and E. Martínez-Salas, "Relevance of RNA structure for the activity of picornavirus IRES elements," *Virus Research*, vol. 139, no. 2, pp. 172–182, 2009.

[9] M. Sáiz, S. Gómez, E. Martínez-Salas, and F. Sobrino, "Deletion or substitution of the aphthovirus 3′ NCR abrogates infectivity and virus replication," *Journal of General Virology*, vol. 82, no. 1, pp. 93–101, 2001.

[10] S. López de Quinto, M. Sáiz, D. de La Morena, F. Sobrino, and E. Martínez-Salas, "IRES-driven translation is stimulated separately by the FMDV 3′-NCR and poly(A) sequences," *Nucleic Acids Research*, vol. 30, no. 20, pp. 4398–4405, 2002.

[11] E. Dobrikova, P. Florez, S. Bradrick, and M. Gromeier, "Activity of a type 1 picornavirus internal ribosomal entry site is determined by sequences within the 3′ nontranslated region," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 15125–15130, 2003.

[12] S. López de Quinto and E. Martínez-Salas, "Involvement of the aphthovirus RNA region located between the two functional AUGs in start codon selection," *Virology*, vol. 255, no. 2, pp. 324–336, 1999.

[13] D. E. Andreev, O. Fernández-Miragall, J. Ramajo, et al., "Differential factor requirement to assemble translation initiation complexes at the alternative start codons of foot-and-mouth disease virus RNA," *RNA*, vol. 13, no. 8, pp. 1366–1374, 2007.

[14] E. Martínez-Salas, "The impact of RNA structure on picornavirus IRES activity," *Trends in Microbiology*, vol. 16, no. 5, pp. 230–237, 2008.

[15] G. J. Belsham, "Divergent picornavirus IRES elements," *Virus Research*, vol. 139, no. 2, pp. 183–192, 2009.

[16] P. J. Lukavsky, "Structure and function of HCV IRES domains," *Virus Research*, vol. 139, no. 2, pp. 166–171, 2009.

[17] J. S. Kieft, "Comparing the three-dimensional structures of Dicistroviridae IGR IRES RNAs with other viral RNA structures," *Virus Research*, vol. 139, no. 2, pp. 148–156, 2009.

[18] S. López de Quinto and E. Martínez-Salas, "Conserved structural motifs located in distal loops of aphthovirus internal ribosome entry site domain 3 are required for internal initiation of translation," *Journal of Virology*, vol. 71, no. 5, pp. 4171–4175, 1997.

[19] M. E. M. Robertson, R. A. Seamons, and G. J. Belsham, "A selection system for functional internal ribosome entry site (IRES) elements: analysis of the requirement for a conserved GNRA tetraloop in the encephalomyocarditis virus IRES," *RNA*, vol. 5, no. 9, pp. 1167–1179, 1999.

[20] O. Fernández-Miragall and E. Martínez-Salas, "Structural organization of a viral IRES depends on the integrity of the GNRA motif," *RNA*, vol. 9, no. 11, pp. 1333–1344, 2003.

[21] O. Fernández-Miragall, R. Ramos, J. Ramajo, and E. Martínez-Salas, "Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation," *RNA*, vol. 12, no. 2, pp. 223–234, 2006.

[22] P. Serrano, J. Gomez, and E. Martínez-Salas, "Characterization of a cyanobacterial RNase P ribozyme recognition motif in the IRES of foot-and-mouth disease virus reveals a unique structural element," *RNA*, vol. 13, no. 6, pp. 849–859, 2007.

[23] P. Serrano, J. Ramajo, and E. Martínez-Salas, "Rescue of internal initiation of translation by RNA complementation provides evidence for a distribution of functions between individual IRES domains," *Virology*, vol. 388, no. 1, pp. 221–229, 2009.

[24] V. G. Kolupaeva, T. V. Pestova, C. U. T. Hellen, and I. N. Shatsky, "Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA," *Journal of Biological Chemistry*, vol. 273, no. 29, pp. 18599–18604, 1998.

[25] S. de Breyne, Y. Yu, A. Unbehaun, T. V. Pestova, and C. U. T. Hellen, "Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 23, pp. 9197–9202, 2009.

[26] S. López de Quinto, E. Lafuente, and E. Martínez-Salas, "IRES interaction with translation initiation factors: functional characterization of novel RNA contacts with eIF3, eIF4B, and eIF4GII," *RNA*, vol. 7, no. 9, pp. 1213–1226, 2001.

[27] S. López de Quinto and E. Martínez-Salas, "Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo," *RNA*, vol. 6, no. 10, pp. 1380–1392, 2000.

[28] A. T. Clark, M. E. M. Robertson, G. L. Conn, and G. J. Belsham, "Conserved nucleotides within the J domain of the encephalomyocarditis virus internal ribosome entry site are required for activity and for interaction with eIF4G," *Journal of Virology*, vol. 77, no. 23, pp. 12441–12449, 2003.

[29] A. J. Dorner, B. L. Semler, R. J. Jackson, R. Hanecak, E. Duprey, and E. Wimmer, "In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate," *Journal of Virology*, vol. 50, no. 2, pp. 507–514, 1984.

[30] J. R. Gebhard and E. Ehrenfeld, "Specific interactions of HeLa cell proteins with proposed translation domains of the poliovirus 5′ noncoding region," *Journal of Virology*, vol. 66, no. 5, pp. 3101–3109, 1992.

[31] B. M. Lunde, C. Moore, and G. Varani, "RNA-binding proteins: modular design for efficient function," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 6, pp. 479–490, 2007.

[32] K. E. Lukong, K. Chang, E. W. Khandjian, and S. Richard, "RNA-binding proteins in human genetic disease," *Trends in Genetics*, vol. 24, no. 8, pp. 416–425, 2008.

[33] E. Martínez-Salas, O. Fernández-Miragall, S. Reigadas, A. Pacheco, and P. Serrano, "Internal ribosome entry site elements in eukaryotic genomes," *Current Genomics*, vol. 5, no. 3, pp. 259–277, 2004.

[34] K. Sawicka, M. Bushell, K. A. Spriggs, and A. E. Willis, "Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein," *Biochemical Society Transactions*, vol. 36, no. 4, pp. 641–647, 2008.

[35] S. Reigadas, A. Pacheco, J. Ramajo, S. López de Quinto, and E. Martínez-Salas, "Specific interference between two unrelated internal ribosome entry site elements impairs translation efficiency," *FEBS Letters*, vol. 579, no. 30, pp. 6803–6808, 2005.

[36] A. Pacheco, S. Reigadas, and E. Martínez-Salas, "Riboproteomic analysis of polypeptides interacting with the internal ribosome-entry site element of foot-and-mouth disease viral RNA," *Proteomics*, vol. 8, no. 22, pp. 4782–4790, 2008.

[37] Y. Yu, H. Ji, J. A. Doudna, and J. A. Leary, "Mass spectrometric analysis of the human 40S ribosomal subunit: native and HCV IRES-bound complexes," *Protein Science*, vol. 14, no. 6, pp. 1438–1446, 2005.

[38] S. Weinlich, S. Hüttelmaier, A. Schierhorn, S.-E. Behrens, A. Ostareck-Lederer, and D. H. Ostareck, "IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3′UTR," *RNA*, vol. 15, no. 8, pp. 1528–1542, 2009.

[39] N. Luz and E. Beck, "A cellular 57 kDa protein binds to two regions of the internal translation initiation site of foot-and-mouth disease virus," *FEBS Letters*, vol. 269, no. 2, pp. 311–314, 1990.

[40] S. K. Jang and E. Wimmer, "Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein," *Genes and Development*, vol. 4, no. 9, pp. 1560–1572, 1990.

[41] A. Borman, M. T. Howell, J. G. Patton, and R. J. Jackson, "The involvement of a spliceosome component in internal initiation of human rhinovirus RNA translation," *Journal of General Virology*, vol. 74, no. 9, pp. 1775–1788, 1993.

[42] V. G. Kolupaeva, C. U. T. Hellen, and I. N. Shatsky, "Structural analysis of the interaction of the pyrimidine tract-binding protein with the internal ribosomal entry site of encephalomyocarditis virus and foot-and-mouth disease virus RNAs," *RNA*, vol. 2, no. 12, pp. 1199–1212, 1996.

[43] P. Kafasla, N. Morgner, T. A. A. Pöyry, S. Curry, C. V. Robinson, and R. J. Jackson, "Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation," *Molecular Cell*, vol. 34, no. 5, pp. 556–568, 2009.

[44] M. R. Conte, T. Grüne, J. Ghuman, et al., "Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold," *EMBO Journal*, vol. 19, no. 12, pp. 3132–3141, 2000.

[45] P. M. Florez, O. M. Sessions, E. J. Wagner, M. Gromeier, and M. A. Garcia-Blanco, "The polypyrimidine tract binding protein is required for efficient picornavirus gene expression and propagation," *Journal of Virology*, vol. 79, no. 10, pp. 6172–6179, 2005.

[46] Y. K. Kim, B. Hahm, and S. K. Jang, "Polypyrimidine tract-binding protein inhibits translation of bip mRNA," *Journal of Molecular Biology*, vol. 304, no. 2, pp. 119–133, 2000.

[47] E. V. Pilipenko, E. G. Viktorova, S. T. Guest, V. I. Agol, and R. P. Roos, "Cell-specific proteins regulate viral RNA translation and virus-induced disease," *EMBO Journal*, vol. 20, no. 23, pp. 6899–6908, 2001.

[48] M. K. Merrill, E. Y. Dobrikova, and M. Gromeier, "Cell-type-specific repression of internal ribosome entry site activity by double-stranded RNA-binding protein 76," *Journal of Virology*, vol. 80, no. 7, pp. 3147–3156, 2006.

[49] M. K. Merrill and M. Gromeier, "The double-stranded RNA binding protein 76:NF45 heterodimer inhibits translation initiation at the rhinovirus type 2 internal ribosome entry site," *Journal of Virology*, vol. 80, no. 14, pp. 6936–6942, 2006.

[50] Y. Song, E. Tzima, K. Ochs, et al., "Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation," *RNA*, vol. 11, no. 12, pp. 1809–1824, 2005.

[51] E. V. Pilipenko, T. V. Pestova, V. G. Kolupaeva, et al., "A cell cycle-dependent protein serves as a template-specific translation initiation factor," *Genes and Development*, vol. 14, no. 16, pp. 2028–2045, 2000.

[52] T. P. Monie, A. J. Perrin, J. R. Birtley, et al., "Structural insights into the transcriptional and translational roles of Ebp1," *EMBO Journal*, vol. 26, no. 17, pp. 3936–3944, 2007.

[53] O. Boussadia, M. Niepmann, L. Créancier, A.-C. Prats, F. Dautry, and H. Jacquemin-Sablon, "Unr is required in vivo for efficient initiation of translation from the internal ribosome entry sites of both rhinovirus and poliovirus," *Journal of Virology*, vol. 77, no. 6, pp. 3353–3359, 2003.

[54] E. C. Anderson, S. L. Hunt, and R. J. Jackson, "Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the 5′ untranslated region," *Journal of General Virology*, vol. 88, no. 11, pp. 3043–3052, 2007.

[55] H. Lu, W. Li, W. S. Noble, D. Payan, and D. C. Anderson, "Riboproteomics of the hepatitis C virus internal ribosomal entry site," *Journal of Proteome Research*, vol. 3, no. 5, pp. 949–957, 2004.

[56] J. R. Evans, S. A. Mitchell, K. A. Spriggs, et al., "Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo," *Oncogene*, vol. 22, no. 39, pp. 8012–8020, 2003.

[57] S. A. Mitchell, K. A. Spriggs, M. J. Coldwell, R. J. Jackson, and A. E. Willis, "The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and Unr," *Molecular Cell*, vol. 11, no. 3, pp. 757–771, 2003.

[58] B. Schepens, S. A. Tinton, Y. Bruynooghe, et al., "A role for hnRNP C1/C2 and Unr in internal initiation of translation during mitosis," *EMBO Journal*, vol. 26, no. 1, pp. 158–169, 2007.

[59] H. Matsui, H. Asou, and T. Inaba, "Cytokines direct the regulation of Bim mRNA stability by heat-shock cognate protein 70," *Molecular Cell*, vol. 25, no. 1, pp. 99–112, 2007.

[60] K. M. Bedard, S. Daijogo, and B. L. Semler, "A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation," *EMBO Journal*, vol. 26, no. 2, pp. 459–467, 2007.

[61] A. Pacheco, S. López de Quinto, J. Ramajo, N. Fernández, and E. Martínez-Salas, "A novel role for Gemin5 in mRNA translation," *Nucleic Acids Research*, vol. 37, no. 2, pp. 582–590, 2009.

[62] D. J. Battle, C.-K. Lau, L. Wan, H. Deng, F. Lotti, and G. Drey-fuss, "The Gemin5 protein of the SMN complex identifies snRNAs," *Molecular Cell*, vol. 23, no. 2, pp. 273–279, 2006.

[63] C. L. Will and R. Lührmann, "Spliceosomal UsnRNP biogenesis, structure and function," *Current Opinion in Cell Biology*, vol. 13, no. 3, pp. 290–301, 2001.

[64] I. Fierro-Monti, S. Mohammed, R. Matthiesen, et al., "Quantitative proteomics identifies Gemin5, a scaffolding protein involved in ribonucleoprotein assembly, as a novel partner for eukaryotic initiation factor 4E," *Journal of Proteome Research*, vol. 5, no. 6, pp. 1367–1378, 2006.

[65] G. Michlewski, J. R. Sanford, and J. F. Cáceres, "The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1," *Molecular Cell*, vol. 30, no. 2, pp. 179–189, 2008.

[66] M. Kiledjian and G. Dreyfuss, "Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box," *EMBO Journal*, vol. 11, no. 7, pp. 2655–2664, 1992.

[67] J. H. Kim, B. Hahm, Y. K. Kim, M. Choi, and S. K. Jang, "Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm," *Journal of Molecular Biology*, vol. 298, no. 3, pp. 395–405, 2000.

[68] A. V. Makeyev and S. A. Liebhaber, "The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms," *RNA*, vol. 8, no. 3, pp. 265–278, 2002.

[69] M. Lynch, L. Chen, M. J. Ravitz, et al., "hnRNP K binds a core polypyrimidine element in the eukaryotic translation initiation factor 4E (eIF4E) promoter, and its regulation of eIF4E contributes to neoplastic transformation," *Molecular and Cellular Biology*, vol. 25, no. 15, pp. 6436–6453, 2005.

[70] P.-T. Lee, P.-C. Liao, W.-C. Chang, and J. T. Tseng, "Epidermal growth factor increases the interaction between nucleolin and heterogeneous nuclear ribonucleoprotein K/poly(C) binding protein 1 complex to regulate the gastrin mRNA turnover," *Molecular Biology of the Cell*, vol. 18, no. 12, pp. 5004–5013, 2007.

[71] B. Collier, B. Gorgoni, C. Loveridge, H. J. Cooke, and N. K. Gray, "The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells," *EMBO Journal*, vol. 24, no. 14, pp. 2656–2666, 2005.

[72] L. B. Blyn, J. S. Towner, B. L. Semler, and E. Ehrenfeld, "Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA," *Journal of Virology*, vol. 71, no. 8, pp. 6243–6246, 1997.

[73] A. V. Gamarnik, N. Böddeker, and R. Andino, "Translation and replication of human rhinovirus type 14 and mengovirus in Xenopus oocytes," *Journal of Virology*, vol. 74, no. 24, pp. 11983–11987, 2000.

[74] K. Choi, J. H. Kim, X. Li, et al., "Identification of cellular proteins enhancing activities of internal ribosomal entry sites by competition with oligodeoxynucleotides," *Nucleic Acids Research*, vol. 32, no. 4, pp. 1308–1317, 2004.

[75] P. Sean, J. H. C. Nguyen, and B. L. Semler, "Altered interactions between stem-loop IV within the 5′ noncoding region of coxsackievirus RNA and poly(rC) binding protein 2: effects on IRES-mediated translation and viral infectivity," *Virology*, vol. 389, no. 1-2, pp. 45–58, 2009.

[76] B. L. Walter, J. H. C. Nguyen, E. Ehrenfeld, and B. L. Semler, "Differential utilization of poly(rC) binding protein 2 in translation directed by picornavirus IRES elements," *RNA*, vol. 5, no. 12, pp. 1570–1585, 1999.

[77] E. Martínez-Salas, S. López de Quinto, R. Ramos, and O. Fernández-Miragall, "IRES elements: features of the RNA structure contributing to their activity," *Biochimie*, vol. 84, no. 8, pp. 755–763, 2002.

[78] R. Perera, S. Daijogo, B. L. Walter, J. H. C. Nguyen, and B. L. Semler, "Cellular protein modification by poliovirus: the two faces of poly(rC)-binding protein," *Journal of Virology*, vol. 81, no. 17, pp. 8919–8932, 2007.

[79] J. Graff, J. Cha, L. B. Blyn, and E. Ehrenfeld, "Interaction of poly(rC) binding protein 2 with the 5′ noncoding region of hepatitis A virus RNA and its effects on translation," *Journal of Virology*, vol. 72, no. 12, pp. 9668–9675, 1998.

[80] C. Candé, N. Vahsen, D. Métivier, et al., "Regulation of cytoplasmic stress granules by apoptosis-inducing factor," *Journal of Cell Science*, vol. 117, no. 19, pp. 4461–4468, 2004.

[81] K. Fujimura, J. Katahira, F. Kano, Y. Yoneda, and M. Murata, "Selective localization of PCBP2 to cytoplasmic processing bodies," *Biochimica et Biophysica Acta*, vol. 1793, no. 5, pp. 878–887, 2009.

[82] F. Parker, F. Maurier, I. Delumeau, et al., "A Ras-GTPase-activating protein SH3-domain-binding protein," *Molecular and Cellular Biology*, vol. 16, no. 6, pp. 2561–2569, 1996.

[83] S. Solomon, Y. Xu, B. Wang, et al., "Distinct structural features of caprin-1 mediate its interaction with G3BP-1 and its induction of phosphorylation of eukaryotic translation initiation factor 2α, entry to cytoplasmic stress granules, and selective interaction with a subset of mRNAs," *Molecular and Cellular Biology*, vol. 27, no. 6, pp. 2324–2342, 2007.

[84] H. Tourrière, K. Chebli, L. Zekri, et al., "The RasGAP-associated endoribonuclease G3BP assembles stress granules," *Journal of Cell Biology*, vol. 160, no. 6, pp. 823–831, 2003.

[85] P. Tingting, F. Caiyun, Y. Zhigang, Y. Pengyuan, and Y. Zhenghong, "Subproteomic analysis of the cellular proteins associated with the 3′ untranslated region of the hepatitis C virus genome in human liver cells," *Biochemical and Biophysical Research Communications*, vol. 347, no. 3, pp. 683–691, 2006.

[86] K. E. Gustin and P. Sarnow, "Effects of poliovirus infection on nucleo-cytoplasmic trafficking and nuclear pore complex composition," *EMBO Journal*, vol. 20, no. 1-2, pp. 240–249, 2001.

[87] H. Ginisty, H. Sicard, B. Roger, and P. Bouvet, "Structure and functions of nucleolin," *Journal of Cell Science*, vol. 112, no. 6, pp. 761–772, 1999.

[88] S. Waggoner and P. Sarnow, "Viral ribonucleoprotein complex formation and nucleolar-cytoplasmic relocalization of nucleolin in poliovirus-infected cells," *Journal of Virology*, vol. 72, no. 8, pp. 6699–6709, 1998.

[89] R. E. Izumi, B. Valdez, R. Banerjee, M. Srivastava, and A. Dasgupta, "Nucleolin stimulates viral internal ribosome entry site-mediated translation," *Virus Research*, vol. 76, no. 1, pp. 17–29, 2001.

[90] J.-Y. Lin, M.-L. Li, and S.-R. Shih, "Far upstream element binding protein 2 interacts with enterovirus 71 internal ribosomal entry site and negatively regulates viral translation," *Nucleic Acids Research*, vol. 37, no. 1, pp. 47–59, 2009.

[91] J. Höck, L. Weinmann, C. Ender, et al., "Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells," *EMBO Reports*, vol. 8, no. 11, pp. 1052–1060, 2007.

[92] M. Landthaler, D. Gaidatzis, A. Rothballer, et al., "Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs," *RNA*, vol. 14, no. 12, pp. 2580–2596, 2008.

[93] K. Tsukiyama-Kohara, N. Iizuka, M. Kohara, and A. Nomoto, "Internal ribosome entry site within hepatitis C virus RNA," *Journal of Virology*, vol. 66, no. 3, pp. 1476–1483, 1992.

[94] C. S. Fraser and J. A. Doudna, "Structural and mechanistic insights into hepatitis C viral translation initiation," *Nature Reviews Microbiology*, vol. 5, no. 1, pp. 29–38, 2007.

[95] C. Wang, S.-Y. Le, N. Ali, and A. Siddiqui, "An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region," *RNA*, vol. 1, no. 5, pp. 526–537, 1995.

[96] G. A. Otto and J. D. Puglisi, "The pathway of HCV IRES-mediated translation initiation," *Cell*, vol. 119, no. 3, pp. 369–380, 2004.

[97] N. Locker, L. E. Easton, and P. J. Lukavsky, "HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly," *EMBO Journal*, vol. 26, no. 3, pp. 795–805, 2007.

[98] J. S. Kieft, K. Zhou, R. Jubin, and J. A. Doudna, "Mechanism of ribosome recruitment by hepatitis C IRES RNA," *RNA*, vol. 7, no. 2, pp. 194–206, 2001.

[99] E. Buratti, S. Tisminetzky, M. Zotti, and F. E. Baralle, "Functional analysis of the interaction between HCV 5′UTR and putative subunits of eukaryotic translation initiation factor eIF3," *Nucleic Acids Research*, vol. 26, no. 13, pp. 3179–3187, 1998.

[100] B. Siridechadilok, C. S. Fraser, R. J. Hall, J. A. Doudna, and E. Nogales, "Molecular biology: structural roles for human translation factor eIF3 in initiation of protein synthesis," *Science*, vol. 310, no. 5753, pp. 1513–1515, 2005.

[101] I. M. Terenin, S. E. Dmitriev, D. E. Andreev, and I. N. Shatsky, "Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2," *Nature Structural and Molecular Biology*, vol. 15, no. 8, pp. 836–841, 2008.

[102] S. Fukushi, M. Okada, J. Stahl, T. Kageyama, F. B. Hoshino, and K. Katayama, "Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus," *Journal of Biological Chemistry*, vol. 276, no. 24, pp. 20824–20826, 2001.

[103] E. Laletina, D. Graifer, A. Malygin, A. Ivanov, I. Shatsky, and G. Karpova, "Proteins surrounding hairpin IIIe of the hepatitis C virus internal ribosome entry site on the human 40S ribosomal subunit," *Nucleic Acids Research*, vol. 34, no. 7, pp. 2027–2036, 2006.

[104] G. A. Otto, P. J. Lukavsky, A. M. Lancaster, P. Sarnow, and J. D. Puglisi, "Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction," *RNA*, vol. 8, no. 7, pp. 913–923, 2002.

[105] H. Ji, C. S. Fraser, Y. Yu, J. Leary, and J. A. Doudna, "Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 49, pp. 16990–16995, 2004.

[106] J. Sengupta, J. Nilsson, R. Gursky, C. M. T. Spahn, P. Nissen, and J. Frank, "Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM," *Nature Structural and Molecular Biology*, vol. 11, no. 10, pp. 957–962, 2004.

[107] O. Isken, M. Baroth, C. W. Grassmann, et al., "Nuclear factors are involved in hepatitis C virus RNA replication," *RNA*, vol. 13, no. 10, pp. 1675–1692, 2007.

[108] C. Romero-López and A. Berzal-Herranz, "A long-range RNA-RNA interaction between the 5′ and 3′ ends of the HCV genome," *RNA*, vol. 15, no. 9, pp. 1740–1752, 2009.

[109] P. Serrano, M. R. Pulido, M. Saiz, and E. Martínez-Salas, "The 3′ end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA-RNA interactions with the 5′ and region," *Journal of General Virology*, vol. 87, no. 10, pp. 3013–3022, 2006.

[110] R. Gosert, K. H. Chang, R. Rijnbrand, M. Yi, D. V. Sangar, and S. M. Lemon, "Transient expression of cellular polypyrimidine-tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites in vivo," *Molecular and Cellular Biology*, vol. 20, no. 5, pp. 1583–1595, 2000.

[111] Y. Song, P. Friebe, E. Tzima, C. Jünemann, R. Bartenschlager, and M. Niepmann, "The hepatitis C virus RNA 3′-untranslated region strongly enhances translation directed by the internal ribosome entry site," *Journal of Virology*, vol. 80, no. 23, pp. 11579–11588, 2006.

[112] J. H. Kim, K. Y. Paek, S. H. Ha, et al., "A cellular RNA-binding protein enhances internal ribosomal entry site-dependent translation through an interaction downstream of the hepatitis C virus polyprotein initiation codon," *Molecular and Cellular Biology*, vol. 24, no. 18, pp. 7878–7890, 2004.

[113] T. Mondal, U. Ray, A. K. Manna, R. Gupta, S. Roy, and S. Das, "Structural determinant of human La protein critical for internal initiation of translation of hepatitis C virus RNA," *Journal of Virology*, vol. 82, no. 23, pp. 11927–11938, 2008.

[114] K. Meerovitch, Y. V. Svitkin, H. S. Lee, et al., "La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate," *Journal of Virology*, vol. 67, no. 7, pp. 3798–3807, 1993.

[115] Y. K. Kim and S. K. Jang, "La protein is required for efficient translation driven by encephalomyocarditis virus internal ribosomal entry site," *Journal of General Virology*, vol. 80, no. 12, pp. 3159–3166, 1999.

[116] S. Cordes, Y. Kusov, T. Heise, and V. Gauss-Müller, "La autoantigen suppresses IRES-dependent translation of the hepatitis A virus," *Biochemical and Biophysical Research Communications*, vol. 368, no. 4, pp. 1014–1019, 2008.

[117] H. M. Liu, H. Aizaki, K. S. Choi, K. Machida, J. J.-H. Ou, and M. M. C. Lai, "SYNCRIP (synaptotagmin-binding, cytoplasmic RNA-interacting protein) is a host factor involved in hepatitis C virus RNA replication," *Virology*, vol. 386, no. 2, pp. 249–256, 2009.

[118] B. Hahm, Y. K. Kim, J. H. Kim, T. Y. Kim, and S. K. Jang, "Heterogeneous nuclear ribonucleoprotein L interacts with the 3′ border of the internal ribosomal entry site of hepatitis C virus," *Journal of Virology*, vol. 72, no. 11, pp. 8782–8788, 1998.

[119] B. Hwang, J. H. Lim, B. Hahm, S. K. Jang, and S.-W. Lee, "hnRNP L is required for the translation mediated by HCV IRES," *Biochemical and Biophysical Research Communications*, vol. 378, no. 3, pp. 584–588, 2009.

[120] M. Majumder, I. Yaman, F. Gaccioli, et al., "The hnRNA-binding proteins hnRNP L and PTB are required for efficient translation of the Cat-1 arginine/lysine transporter mRNA during amino acid starvation," *Molecular and Cellular Biology*, vol. 29, no. 10, pp. 2899–2912, 2009.

[121] K. Y. Paek, C. S. Kim, S. M. Park, J. H. Kim, and S. K. Jang, "RNA-binding protein hnRNP D modulates internal ribosome entry site-dependent translation of hepatitis C Virus RNA," *Journal of Virology*, vol. 82, no. 24, pp. 12082–12093, 2008.

[122] J.-Y. Lin, S.-R. Shih, M. Pan, et al., "hnRNP A1 interacts with the 3′ untranslated regions of enterovirus 71 and Sindbis virus RNA and is required for viral replication," *Journal of Virology*, vol. 83, no. 12, pp. 6106–6114, 2009.

[123] S. Bonnal, F. Pileur, C. Orsini, et al., "Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA," *Journal of Biological Chemistry*, vol. 280, no. 6, pp. 4144–4153, 2005.

[124] A. Cammas, F. Pileur, S. Bonnal, et al., "Cytoplasmic relocalization of heterogeneous nuclear ribonucleoprotein A1 controls translation initiation of specific mRNAs," *Molecular Biology of the Cell*, vol. 18, no. 12, pp. 5048–5059, 2007.

[125] A. Kukalev, Y. Nord, C. Palmberg, T. Bergman, and P. Percipalle, "Actin and hnRNP U cooperate for productive transcription by RNA polymerase II," *Nature Structural and Molecular Biology*, vol. 12, no. 3, pp. 238–244, 2005.

[126] P. Percipalle, A. Jonsson, D. Nashchekin, et al., "Nuclear actin is associated with a specific subset of hnRNP A/B-type proteins," *Nucleic Acids Research*, vol. 30, no. 8, pp. 1725–1734, 2002.

[127] D. E. Schultz, C. C. Hardin, and S. M. Lemon, "Specific interaction of glyceraldehyde 3-phosphate dehydrogenase with the 5′-nontranslated RNA of hepatitis a virus," *Journal of Biological Chemistry*, vol. 271, no. 24, pp. 14134–14142, 1996.

[128] M. Yi, D. E. Schultz, and S. M. Lemon, "Functional significance of the interaction of hepatitis A virus RNA with glyceraldehyde 3-phosphate dehydrogenase (GAPDH): opposing effects of GAPDH and polypyrimidine tract binding protein on internal ribosome entry site function," *Journal of Virology*, vol. 74, no. 14, pp. 6459–6468, 2000.

[129] R. Mukhopadhyay, J. Jia, A. Arif, P. S. Ray, and P. L. Fox, "The GAIT system: a gatekeeper of inflammatory gene expression," *Trends in Biochemical Sciences*, vol. 34, no. 7, pp. 324–331, 2009.

[130] F. V. Fuller-Pace, "DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation," *Nucleic Acids Research*, vol. 34, no. 15, pp. 4206–4215, 2006.

[131] T. R. Hartman, S. Qian, C. Bolinger, S. Fernandez, D. R. Schoenberg, and K. Boris-Lawrie, "RNA helicase A is necessary for translation of selected messenger RNAs," *Nature Structural and Molecular Biology*, vol. 13, no. 6, pp. 509–516, 2006.

[132] H.-C. Chen, W.-C. Lin, Y.-G. Tsay, S.-C. Lee, and C.-J. Chang, "An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K," *Journal of Biological Chemistry*, vol. 277, no. 43, pp. 40403–40409, 2002.

[133] M. Rodríguez Pulido, P. Serrano, M. Sáiz, and E. Martínez-Salas, "Foot-and-mouth disease virus infection induces proteolytic cleavage of PTB, eIF3a,b, and PABP RNA-binding proteins," *Virology*, vol. 364, no. 2, pp. 466–474, 2007.

[134] J. P. White, A. M. Cardenas, W. E. Marissen, and R. E. Lloyd, "Inhibition of cytoplasmic mRNA stress granule formation by a viral proteinase," *Cell Host and Microbe*, vol. 2, no. 5, pp. 295–305, 2007.

[135] V. Guerniou, R. Gillet, F. Berrée, B. Carboni, and B. Felden, "Targeted inhibition of the hepatitis C internal ribosomal entry site genomic RNA with oligonucleotide conjugates," *Nucleic Acids Research*, vol. 35, no. 20, pp. 6778–6787, 2007.

[136] C. Romero-López, R. Díaz-González, A. Barroso-del Jesus, and A. Berzal-Herranz, "Inhibition of hepatitis C virus replication and internal ribosome entry site-dependent translation by an RNA molecule," *Journal of General Virology*, vol. 90, no. 7, pp. 1659–1669, 2009.

[137] M. F. Rosas, E. Martínez-Salas, and F. Sobrino, "Stable expression of antisense RNAs targeted to the 5′ non-coding region confers heterotypic inhibition to foot-and-mouth disease virus infection," *Journal of General Virology*, vol. 84, no. 2, pp. 393–402, 2003.

[138] A. Vagnozzi, D. A. Stein, P. L. Iversen, and E. Rieder, "Inhibition of foot-and-mouth disease virus infections in cell cultures with antisense morpholino oligomers," *Journal of Virology*, vol. 81, no. 21, pp. 11669–11680, 2007.

Review Article

The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of Any Given Proteome

Vladimir N. Uversky[1, 2, 3]

[1] *Institute for Intrinsically Disordered Protein Research, The Center for Computational Biology and Bioinformatics,*
Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[2] *Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia*
[3] *Molecular Kinetics Inc., Indianapolis, IN 46268, USA*

Correspondence should be addressed to Vladimir N. Uversky, vuversky@iupui.edu

Contrarily to the general believe, many biologically active proteins lack stable tertiary and/or secondary structure under physiological conditions in vitro. These intrinsically disordered proteins (IDPs) are highly abundant in nature and many of them are associated with various human diseases. The functional repertoire of IDPs complements the functions of ordered proteins. Since IDPs constitute a significant portion of any given proteome, they can be combined in an unfoldome; which is a portion of the proteome including all IDPs (also known as natively unfolded proteins, therefore, unfoldome), and describing their functions, structures, interactions, evolution, and so forth. Amino acid sequence and compositions of IDPs are very different from those of ordered proteins, making possible reliable identification of IDPs at the proteome level by various computational means. Furthermore, IDPs possess a number of unique structural properties and are characterized by a peculiar conformational behavior, including their high stability against low pH and high temperature and their structural indifference toward the unfolding by strong denaturants. These peculiarities were shown to be useful for elaboration of the experimental techniques for the large-scale identification of IDPs in various organisms. Some of the computational and experimental tools for the unfoldome discovery are discussed in this review.

1. Introducing Unfoldomes and Unfoldomics

Proteins are the major components of the living cell. They play crucial roles in the maintenance of life and protein dysfunctions may cause development of various pathological conditions. Although for a very long time it has been believed that the specific functionality of a given protein is predetermined by its unique 3D structure [1, 2], it is recognized now that the fate of any given polypeptide chain is determined by the peculiarities of its amino acid sequence. In fact, Figure 1 shows that although many proteins are indeed predisposed to fold into unique structures which evolved to possess unique biological functions, some proteins can misfold either spontaneously or due to the mutations and other genetic alterations, problematic processing or posttranslational modifications, or due to the exposure to harmful environmental conditions. Such misfolding is now considered as a crucial early step in the development of various protein conformation diseases [3]. Finally, for more than five decades, researchers have been discovering individual proteins that possess no definite ordered 3D structure but still play important biological roles. The discovery rate for such proteins has been increasing continually and has become especially rapid during the last decade [4]. Such proteins are widely known as intrinsically disordered proteins (IDPs) among other names and are characterized by the lack of a well-defined 3D structure under physiological conditions. The discovery and characterization of these proteins is becoming one of the fastest growing areas of protein science and it is recognized now that many such proteins with no unique structure have important biological functions [4–16]. Structural flexibility and plasticity originating from the lack of a definite-ordered 3D structure are believed to represent a major functional advantage for these proteins,
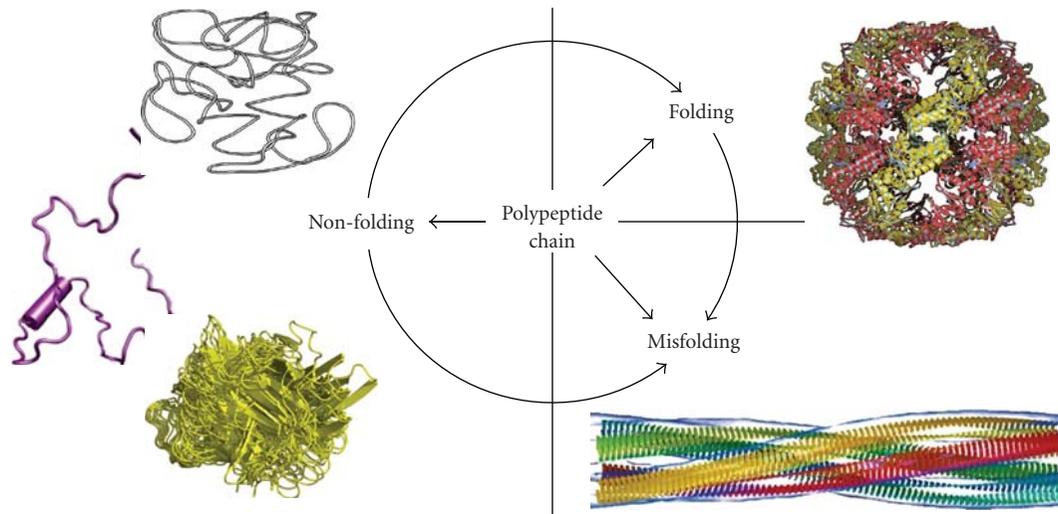
FIGURE 1: *Fate of a polypeptide chain. Left.* Three structures representing typical IDPs with different disorderedness levels (from top to bottom): native coil, native premolten globule, and native molten globule. *Right.* Top structure illustrates a well-folded protein, whereas the bottom structure represents one of the products of protein misfolding—a molecular model of the compact, 4-protofilament insulin fibril (http://people.cryst.bbk.ac.uk/~ubcg16z/amyloid/insmod.jpg).

enabling them to interact with a broad range of binding partners including other proteins, membranes, nucleic acids, and various small molecules [17–19].

The functions attributed to IDPs were grouped into four broad classes: (1) molecular recognition; (2) molecular assembly; (3) protein modification; and (4) entropic chain activities [5, 6]. IDPs are often involved in regulatory/signaling interactions with multiple partners that require high specificity and low affinity [7, 20]. Some illustrative biological activities of IDPs include regulation of cell division, transcription and translation, signal transduction, protein phosphorylation, storage of small molecules, chaperone action, and regulation of the self-assembly of large multiprotein complexes such as the ribosome [4–10, 13–16, 20–27]. The crucial role of IDPs in signaling is further confirmed by the fact that eukaryotic proteomes, with their extensively developed interaction networks, are highly enriched in IDPs, relative to bacteria and archaea (see below, [28–30]). Recently, application of a novel data mining tool to over 200 000 proteins from Swiss-Prot database revealed that many protein functions are associated with long disordered regions [13–15]. In fact, of the 711 Swiss-Prot functional keywords that were associated with at least 20 proteins, 262 were found to be strongly positively correlated with long intrinsically disordered regions (IDRs), whereas 302 were strongly negatively correlated with such regions [13–15]. Therefore, the functional diversity provided by disordered regions complements functions of ordered protein regions.

Although unbound IDPs are disordered in solution, they often perform their biological functions by binding to their specific partners. This binding involves a disorder-to-order transition in which IDPs adopt a highly structured conformation upon binding to their biological partners [31–38]. In this way, IDPs play diverse roles in regulating the function of their binding partners and in promoting the assembly of supramolecular complexes. Furthermore, because sites within their polypeptide chains are highly accessible, IDPs can undergo extensive posttranslational modifications, such as phosphorylation, acetylation, and/or ubiquitination (sumoylation, etc.), allowing for modulation of their biological activity or function. Intriguingly, IDPs were shown to be highly abundant in various diseases, giving rise to the "disorder in disorders" or D^2 concept which generally summarizes work in this area [39].

As the number of IDPs and IDRs in various proteomes is very large (e.g., for mammals, ~75% of their signaling proteins are predicted to contain long disordered regions (>30 residues), about half of their total proteins are predicted to contain such long disordered regions, and ~25% of their proteins are predicted to be fully disordered), and because IDPs and IDRs have amazing structural variability and possess a very wide variety of functions, the unfoldome and unfoldomics concepts were recently introduced [4, 40, 41]. The use of the suffix "-ome" has a long history while "-omics" is much more recent. The Oxford English Dictionary (OED) attributes "genome" to Hans Winkler from his 1920 work [42]. While the OED suggests that "genome" arose as a portmanteau of "gene" and "chromosome," this does not seem to be supported by literature. Instead, Lederberg and McCray suggest that as a botanist, Winkler must have been familiar with terms such as biome (a biological community), rhizome (a root system), and phyllome (the leaves covering a tree) among others, all of which were in use well before 1920 and all of which signify the collectivity of the units involved [43]. Thus, "ome" implies the complete set of the objects in question, with genome signifying the set of genes of an organism. By changing the "e" in "-ome" to "-ics," the new word is created that indicates the scientific study of the "-ome" in question. For example, officially the change of "genome" to "genomics" occurred in 1987, when a journal

by this name was founded by Victor Lederberg and McCray [43].

Many additional conversions from –ome to –omics have subsequently occurred and a large number of "-omes" have been accepted in biology, including but not limited to the following: genome, proteome, interactome, metabolome, transcriptome, diseasome, toxicogenome, nutrigenome, cytome, oncoproteome, epitome, and glycome, and so forth. For a more complete list, the reader is directed to http://omics.org/.

Overall, the suffixes –ome and –omics imply a new layer of knowledge, especially when a scientist is dealing with the data produced by the large-scale studies, including the high-throughput experiments and the computational/bioinformatics analyses of the large datasets. The unfoldome and unfoldomics concepts are built on the ideas given above [44]. Unfoldome is attributed to a portion of proteome which includes a set of IDPs (also known as natively unfolded proteins, therefore, unfoldome). The term unfoldome is also used to cover segments or regions of proteins that remain unfolded in the functional state. Unfoldomics is the field that focuses on the unfoldome. It considers not only the identities of the set of proteins and protein regions in the unfoldome of a given organism but also their functions, structures, interactions, evolution, and so forth [44].

It is clearly recognized now that the disorderedness is linked to the peculiarities of amino acid sequences, as IDPs/IDRs exhibit low sequence complexity and are generally enriched in polar and charged residues and are depleted of hydrophobic residues (other than proline). These features are consistent with their inability to fold into globular structures and form the basis of computational tools for disorder prediction [8, 10, 45–48]. These same computational tools can also be utilized for the large-scale discovery of IDPs in various proteomes (see below).

Being characterized by specific (and somewhat unique) amino acid sequences, IDPs possess a number of very distinctive structural properties that can be implemented for their discovery. This includes but is not limited to sensitivity to proteolysis [49], aberrant migration during SDS-PAGE [50], insensitivity to denaturing conditions [51], as well as definitive disorder characteristics visualized by CD spectropolarimetry, NMR spectroscopy, small-angle X-ray scattering, hydrodynamic measurement, fluorescence, as well as Raman and infrared spectroscopies [52, 53]. Structurally, intrinsically disordered proteins range from completely unstructured polypeptides to extended partially structured forms to compact disordered ensembles containing substantial secondary structure [4, 8, 9, 23, 54]. Many proteins contain mixtures of ordered and disordered regions. Extended IDPs are known to possess the atypical conformational behavior (such as "turn out" response to acidic pH and high temperature and insensitivity to high concentrations of strong denaturants), which is determined by the peculiarities of their amino acid sequences and the lack of ordered 3D structure [55]. These unique structural features of extended IDPs and their specific conformational behavior were shown to be useful in elaboration the

experimental techniques for the large-scale identification of these important members of the protein kingdom. Three related methods were introduced: a method based on the finding that many proteins that fail to precipitate during perchloric acid or trichloroacetic acid treatment were IDPs [40]; a method utilizing the fact that IDPs possessed high resistance toward the aggregation induced by heat treatment [40, 56, 57]; and a method based on the heat treatment coupled with a novel 2D gel methodology to identify IDPs in cell extracts [56]. It is anticipated that these methodologies, combined with highly sensitive mass spectrometry-based techniques, can be used for the detection and functional characterization of IDPs in various proteomes. Some of the computational and experimental tools for the unfoldome discovery are discussed below in more details.

2. Computational Tools for Uncovering the Unfoldomes

2.1. Some Basic Principles of Disorder Prediction from Amino Acid Sequence. One of the key arguments about the existence and distinctiveness of IDPs came from various computational analyses. Historically, already at the early stages of the field, simple statistical comparisons of amino acid compositions and sequence complexity indicated that disordered and ordered regions are highly different to a significant degree [45, 58–60]. These sequence biases were then exploited to predict disordered regions or wholly disordered proteins with relatively high accuracy and to make crucial estimates about the commonness of disordered proteins in the three kingdoms of life [28, 45, 61].

Similar to the "normal" foldable proteins whose correct folding into the rigid biologically active conformation is determined by amino acid sequence, the absence of rigid structure in the "nontraditional" nonfoldable IDPs is encoded in the specific features of their amino acid sequences. In fact, some of the ID proteins have been discovered due their unusual amino acid sequence compositions and the absence of regular structure in these proteins has been explained by the specific features of their amino acid sequences including the presence of numerous uncompensated charged groups (often negative); that is, a large net charge at neutral pH, arising from the extreme pI values in such proteins [62–64], and a low content of hydrophobic amino acid residues [62, 64]. Interestingly, the first predictor of intrinsic disorder was developed by R.J.P. Williams based on the abnormally high charge/hydrophobic ratio for the two ID proteins; that is, using the same set of attributes, large net charge and low overall hydrophobicity [65]. Although this predictor was used to separate just two ID proteins from a small set of ordered proteins, this paper is significant as being the first indication that ID proteins have amino acid compositions that differ substantially from those of proteins with 3D structure.

Later, this approach was re-invented in a form of charge-hydropathy plot [45]. To this end, 275 natively folded and 91 natively unfolded proteins (i.e., proteins which at physiologic conditions have been reported to have the NMR chemical

shifts of a random-coil, and/or lack significant ordered secondary structure (as determined by CD or FTIR), and/or show hydrodynamic dimensions close to those typical of an unfolded polypeptide chain) have been assembled from the literature searches. From the comparison of these datasets it has been concluded that the combination of low mean hydrophobicity and relatively high net charge represents an important prerequisite for the absence of compact structure in proteins under physiological conditions. This observation was used to develop a charge-hydropathy (CH) plot method of analysis that distinguishes ordered and disordered proteins based only on their net charges and hydropathy [45]. Figure 2(a) represents the original CH-plot and shows that natively unfolded proteins are specifically localized within a unique region of CH phase space. Furthermore, ID and ordered proteins can be separated by a linear boundary, above which a polypeptide chain with a given mean net charge will most probably be unfolded [45]. From the physical viewpoint, such a combination of low hydrophobicity with high net charge as a prerequisite for intrinsic unfoldedness makes perfect sense: high net charge leads to charge-charge repulsion, and low hydrophobicity means less driving force for protein compaction. In other words, these features are characteristic for ID proteins with the coil-like (or close to coil-like) structures. Obviously, such highly disordered proteins represent only a small subset of the ID protein realm.

More detailed analysis was elaborated to gain additional information on the compositional difference between ordered and ID proteins. Comparison of a nonredundant set of ordered proteins with several datasets of disorder (where proteins were grouped based on different techniques, X-ray crystallography, NMR and CD, used to identify disorder) revealed that disordered regions share at least some common sequence features over many proteins [66, 67]. These differences in amino acid compositions are visualized in Figure 2(b). Here, the relative content of each amino acid in a given disordered dataset has been expressed as (Disordered–Ordered)/(Ordered). Thus, negative peaks correspond the amino acids in which the disordered segments are depleted compared with the ordered ones, and positive peaks indicate the amino acids in which ID regions are enriched [8]. The arrangement of the amino acids from least to most flexible was based on the scale established by Vihinen et al. [68]. This scale was defined by the average residue B-factors of the backbone atoms for 92 unrelated proteins. Figure 2(b) shows that the disordered proteins are significantly depleted in bulky hydrophobic (Ile, Leu, and Val) and aromatic amino acid residues (Trp, Tyr, and Phe), which would normally form the hydrophobic core of a folded globular protein, and also possess low content of Cys and Asn residues. The depletion of ID protein in Cys is also crucial as this amino acid residue is known to have a significant contribution to the protein conformation stability via the disulfide bond formation or being involved in coordination of different prosthetic groups. These depleted residues, Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn were proposed to be called order-promoting amino acids. On the other hand, ID proteins were shown to be substantially enriched in polar,

(a)

(b)
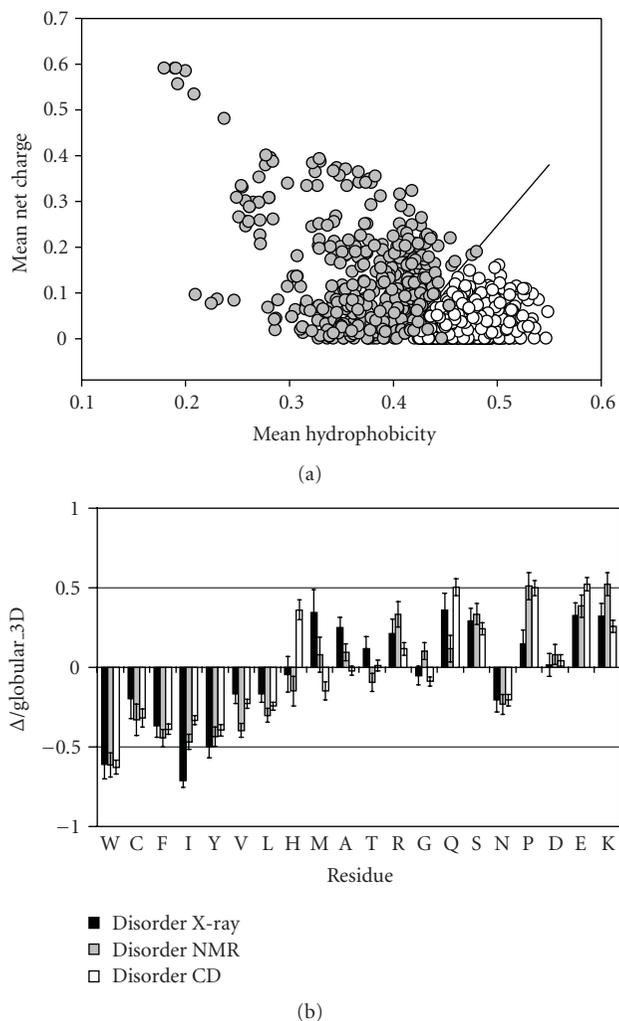
FIGURE 2: *Peculiarities of amino acid composition of ID proteins.* (a) Comparison of the mean net charge and the mean hydrophobicity for a set of 275 ordered (open circles) and 91 natively unfolded proteins (gray circles). The solid line represents the border between extended IDPs and ordered proteins (see text). (b) Order/disorder composition profile. Comparisons of amino acid compositions of ordered proteins with each of three databases of disordered proteins. The ordinates are (%amino acid in disordered dataset − %amino acid in ordered dataset)/(%amino acid in ordered dataset) = Δ/globular_3D. Names of each database indicate how the disordered regions were identified. Negative values indicate that the disordered database has less than order, positive indicates more than order.

disorder-promoting, amino acids: Ala, Arg, Gly, Gln, Ser, Glu, and Lys and also in the hydrophobic, but structure-braking Pro [8, 69, 70]. Note that these biases in the amino acid compositions of ID proteins are also consistent with the low overall hydrophobicity and high net charge characteristic of the natively unfolded proteins [45].

In addition to amino-acid composition, the disordered segments have also been compared with the ordered ones by various attributes such as hydropathy, net charge, flexibility index, helix propensities, strand propensities, and

compositions for groups of amino acids such as W + Y + F (aromaticity). As a result, 265 property-based attribute scales [69] and more than 6000 composition-based attributes (e.g., all possible combinations having one to four amino acids in the group) have been compared [71]. It has been established that ten of these attributes, including 14 Å contact number, hydropathy, flexibility, β-sheet propensity, coordination number, R+E+S+P, bulkiness, C+F+Y+W, volume, and net charge, provide fairly good discrimination between order and disorder [8]. Later, it has been shown that not only the sequence compositions of ordered and disordered regions were different but also that disordered regions of various lengths were diverse as well. In particular, four classes of protein regions were compared: (a) low B-factor ordered regions; (b) high B-factor ordered regions; (c) short disordered regions; and (d) long disordered regions [72]. The four types of regions were shown to have distinct sequence and physicochemical characteristics, with short disordered regions and high B-factor regions being the two closest groups. Furthermore, each of these two groups was closer to the long disordered regions than to the rigid ordered regions. In summary, the analysis of sequence and comparison of their various physicochemical properties indicated that all sets were mutually different [72]. For example, the short disordered and high B-factor regions were shown to be more negatively charged, while long disordered regions were either positively or negatively charged, but on average nearly neutral [72].

As the amino acid sequences of the IDPs and IDRs differ dramatically from those of the ordered proteins and regions, these amino acid sequence differences were used to develop various predictors of intrinsic disorder. As it has been already mentioned, based on a very small number of proteins, Williams suggested an approach for using amino acid sequence for identifying proteins that form random coils rather than globular structures [65], but this approach was never carefully tested. Later, the first well-tested predictors of IDPs were independently published [45, 58].

Protein disorder is a multifaced phenomenon; that is, disordered proteins, being mobile, flexible, and dynamic, might have very different structural features, which range from collapsed molten globule-like conformation to extended coil-like state. It has been suggested that just as an ordered protein is comprised of different types of secondary structure (α-helices, β-strands, β-turns, 3_{10}-helices, and others), ID protein can also be made up of distinguishable types of disorder [73]. To check this hypothesis, a partitioning algorithm based on the differential prediction accuracies has been developed [73]. This algorithm used the notion that a specialized predictor built on a given disorder flavor should have significantly higher same-flavor accuracy than other-flavor predictors or than a global predictor applied to the same given flavor. Application of this partitioning algorithm to known disordered proteins identified three distinctive "flavors" of disorder, arbitrarily called V, C, and S [73]. Importantly, the flavor-specific disordered proteins have been shown to be distinguishable not only by their amino acid compositions but also by disordered sequence locations, and biological functions. Based on these observations, it was proposed that specific flavor-function relationships do exist and thus it is possible (in principle) to identify the functions of disordered regions from their amino acid sequences alone, without any need for specific structural knowledge [73].

Since then, numerous researchers have designed many algorithms to predict disordered proteins utilizing specific biochemical properties and biased amino acid compositions of IDPs. Various prediction ideas and different computing techniques have been utilized. Many of these predictors, including PONDRs [58, 70, 74–77], FoldIndex [78], Glob-Plot [79], DisEMBL [80], DISOPRED and DISOPRED2 [29, 81–83], IUPred [84], FoldUnfold [85], RONN [86], DisPSSMP [87], DisPSSMP2 [88], Spritz [89], and PrDOS [90], and so forth, can be accessed via public servers and evaluate intrinsic disorder on a per-residue basis. Since the first predictors were published, more than 50 predictors of disorder have been developed [91].

It is important to remember that comparing and combining several predictors on an individual protein of interest or on a protein dataset can provide additional insight regarding the predicted disorder if any exists. This is illustrated by a study where two distinct methods for using amino acid sequences to predict which proteins are likely to be mostly disordered, cumulative distribution function (CDF) analysis and charge-hydropathy (CH) plot, have been compared [30]. CDF is based on the PONDR VLXT predictor, which predicts the order-disorder class for every residue in a protein [28, 30]. CDF curves for PONDR VLXT predictions begin at the point with coordinates (0,0) and end at the point with coordinates (1,1) because PONDR VLXT predictions are defined only in the range (0,1) with values less than 0.5 indicating a propensity for order and values greater than or equal to 0.5 indicating a propensity for disorder. The optimal boundary that provided the most accurate order-disorder classification was determined and it has been shown that seven boundary points located in the 12th through 18th bin provided the optimal separation of the ordered and disordered protein sets [30]. For CDF analysis, order-disorder classification is based on whether a CDF curve is above or below a majority of boundary points [30]. In summary, CDF analysis summarizes the per-residue predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores. The other method of order-disorder classification is charge-hydropathy plots [45], in which ordered and disordered proteins being plotted in charge-hydropathy space can be separated to a significant degree by a linear boundary. It has been established that CDF analysis predicts a much higher frequency of disorder in sequence databases than CH-plot discrimination [30]. However, the vast majority of disordered proteins predicted by charge-hydropathy discrimination were also predicted by CDF analysis. These findings are not a big surprise, as CH-plot analysis discriminates protein using only two attributes, mean net charge and mean hydrophobicity, whereas PONDR VLXT (and consequently CDF) is a neural network, which is a nonlinear classifier, trained to distinguish order and disorder based on a relatively large feature space (including average coordination number, amino acid compositions

(aromatic and charged residues), and net charge). Thus, CH feature space can be considered as a subset of PONDR VLXT feature space [30]. Importantly, these findings may be physically interpretable in terms of different types of disorder, collapsed (molten globule-like) and extended (premolten globule- and coil-like). Under this consideration, the CH-plot classification discriminates proteins with the extended disorder from a set of globular conformations (molten globule-like or rigid well-structured proteins) and proteins predicted to be disordered by the CH-plot approach are likely to belong to the extended disorder class. On the other hand, PONDR-based approaches can discriminate all disordered conformations (coil-like, premolten globules, and molten globules) from rigid well-folded proteins, suggesting that CH classification is roughly a subset of PONDR VL-XT, in both predictions of disorder and feature space [30]. Based on this reasoning, several interesting conclusion have been made. It has been suggested that if a protein is predicted to be disordered by both CH and CDF, then, it is likely to be in the extended disorder class. However, a protein predicted to be disordered by CDF but predicted to be ordered by CH-plot might have properties consistent with a dynamic, collapsed chain; that is, it is likely to be in the native molten globule class. Finally, proteins predicted to be ordered by both algorithms are of course likely to be in the well-structured class [30].

2.2. Estimation of Commonness of Disorder in Various Proteomes.

The first application of the disorder predictors was the evaluation of the commonness of protein disorder in the Swiss-Prot database [61]. This analysis revealed that 25% of proteins in Swiss-Prot had predicted ID regions longer than 40 consecutive residues and that at least 11% of residues in Swiss-Prot were likely to be disordered. Given the existence of a few dozen experimentally characterized disordered regions at the time, this work had significant influence on the recognition of the importance of studying disordered proteins [61].

Next, both PONDR VLXT [28] and 3 Flavor PONDR predictors [73] were used to estimate the amount of disorder in various genomes. The predictions counted only disordered regions greater than forty residues in length, which has a false-positive rate of fewer than 400 residues out of 100 000. The result clearly showed that disorder increases from bacteria to archaea to eukaryota with over half of the eukaryotic proteins containing predicted disordered regions [28, 73]. One explanation for this trend is a change in the cellular requirements for certain protein functions, particularly cellular signaling. In support of this hypothesis, PONDR analysis of a eukaryotic signal protein database indicates that the majority of known signal transduction proteins are predicted to contain significant regions of disorder [25].

Ward et al. [29] have refined and systematized such an analysis and concluded that the fraction of proteins containing disordered regions of 30 residues or longer (predicted using DISOPRED) were 2% in archaea, 4% in bacteria, and 33% in eukarya. In addition, a complete

functional analysis of the yeast proteome with respect to the three Gene Ontology (GO) categories was performed. In terms of molecular function, transcription, kinase, nucleic acid, and protein-binding activity were the most distinctive signatures of disordered proteins. The most overrepresented GO terms characteristic for the biological process category were transposition, development, morphogenesis, protein phosphorylation, regulation, transcription, and signal transduction. Finally, with respect to cellular component, it appeared that nuclear proteins were significantly enriched in disorder, while terms such as membrane, cytosol, mitochondrion, and cytoplasm were distinctively overrepresented in ordered proteins [29].

Application of the CH-CDF analysis to various proteomes revealed that CDF analysis predicts about 2-fold higher frequency of disorder in sequence databases than CH-plot classification suggesting that approximately half of disordered proteins in different proteomes possess extended disorder, whereas another half represents proteins with the collapsed disorder [30]. Furthermore, the consensus CDF-CH method showed that approximately 4.5% of *Yersinia pestis*, 5% of *Escherichia coli* K12, 6% of *Archaeoglobus fulgidus*, 8% of *Methanobacterium thermoautotrophicum*, 23% of *Arabidopsis thaliana*, and 28% of *Mus musculus* proteins are wholly disordered [30].

As mentioned above, the CH-plot, being a linear classifier, takes into account only two parameters of the particular sequence—charge and hydropathy, whereas CDF analysis is dependent upon the output of the PONDR VLXT predictor, a nonlinear neural network classifier, which was trained to distinguish order and disorder based on a significantly larger feature space that explicitly includes net charge and hydropathy [30]. According to these methodological differences, CH-plot analysis is predisposed to discriminate proteins with substantial amounts of extended disorder (random coils and premolten globules) from proteins with globular conformations (molten globule-like and rigid well-structured proteins). On the other hand, PONDR-based CDF analysis may discriminate all disordered conformations including molten globules from rigid well-folded proteins [30].

This difference in the sensitivity of predictors to different levels of overall disorderedness was utilized in CDF-CH-plot analysis, which allows the ordered and disordered proteins separation in the CH-CDF phase space [92]. In this approach, each spot corresponds to a single protein and its coordinates are calculated as a distance of this protein from the boundary in the corresponding CH-plot (Y-coordinate) and an averaged distance of the corresponding CDF curve from the boundary (X-coordinate). Positive and negative Y values correspond to proteins which, according to CH-plot analysis, are predicted to be natively unfolded or compact, respectively. Whereas positive and negative X values are attributed to proteins that, by the CDF analysis, are predicted to be ordered or intrinsically disordered, respectively. Therefore, this plot has four quadrants: $(-,-)$ quadrant, which contains proteins predicted to be disordered by CDF, but compact by CH-plot (i.e., proteins with molten globule-like properties); $(-,+)$ quadrant that includes proteins predicted
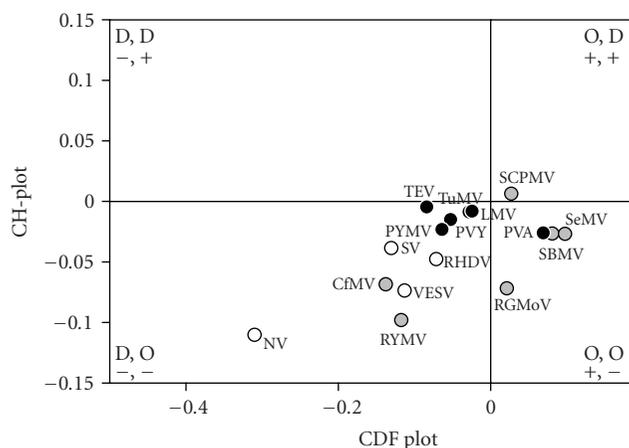
FIGURE 3: *CH-CDF analysis of the genome-linked proteins VPgs from various viruses*. Comparison of the results of PONDR CDF and CH-plot analyses for whole protein order-disorder via distributions of VPgs within the CH-CDF phase space. The protein analyzed by this approach are VPgs from *Sobemovirus* (Rice yellow mottle virus (RYMV), Cocksfoot mottle virus (CoMV), Ryegrass mottle virus (RGMoV), Southern bean mosaic virus (SBMV), Southern cowpea mosaic virus (SCPMV), Sesbania mottle virus (SeMV)), Potyvirus (Lettuce mosaic virus (LMV), Potato virus Y (PVY), Potato virus A (PVA), Tobacco etch virus (TEV), Turnip mosaic virus (TuMV), Bean yellow mosaic virus (BYMV)), and *Caliciviridae* (Rabbit hemorrhabic disease virus (RHDV), Vesicular exanthema of swine virus (VESV), Man Sapporo virus Manchester virus (SV), and Norwalk virus (NV)).

to be disordered by both methods (i.e., random coils and pre-molten globules); (+,−) quadrant which contains ordered proteins; and (+,+) quadrant including proteins which are predicted to be disordered by CH-plot, but ordered by the CDF analysis [92]. Application of such a combined CH-CDF analysis to mice proteins revealed that ~12% mice proteins are likely to belong to the class of extended IDPs (native coils and native premolten globules), whereas ~30% proteins in mouse genome are potential native molten globules.

Recently, the disorderedness of genome-linked proteins VPg from various viruses were evaluated by a combined CH-CDF analysis [93, 94]. The genome-linked protein VPg of Potato virus A (PVA; genus Potyvirus) has essential functions in all critical steps of PVA infection, that is, replication, movement, and virulence. The structural analysis of the recombinant PVA VPg revealed this protein possesses many properties of a native molten globule [93]. In a follow-up study, it has been shown that although VPgs from various viruses (from *Sobemovirus, Potyvirus*, and *Caliciviridae* genera) are highly diverse in size, sequence, and function, many of them were predicted to contain long disordered domains [94]. Figure 3 summarizes these findings by showing the localization of several VPgs within the CH-CDF phase space. Each spot represents a single VPg whose coordinates were calculated as a distance of this protein from the boundary in the corresponding charge-hydropathy plot (CH-plot, Y-coordinate) and an averaged distance of the corresponding cumulative distribution function (CDF) curve from the

boundary (X-coordinate). Figure 3 clearly shows that many VPgs are expected to behave as native molten globules [94].

Several recent reviews summarized the current state of the art in the field of IDP predictions and represent a useful overview of the prediction methods highlighting their advantages and drawbacks [46, 91]. Concluding, the considered-above studies clearly showed that IDPs (both collapsed and extended) are highly abundant in nature and can be reliably identified by various computational means.

2.3. Discovering the Disease-Related Unfoldomes. Misfolding (the failure of a specific peptide or protein to adopt its functional conformational state) and related dysfunction of many proteins were considered as a major cause for the development of different pathological conditions. Such misfolding and dysfunction can originate from point mutation(s) or result from an exposure to internal or external toxins, impaired posttranslational modifications (phosphorylation, advanced glycation, deamidation, racemization, etc.), an increased probability of degradation, impaired trafficking, lost binding partners, or oxidative damage. All these factors can act independently or in association with one another.

Although the formation of various aggregates represents the most visible consequence of protein misfolding and although these aggregates form the basis for the development of various protein deposition diseases, pathogenesis of many more human diseases does not depend on aggregation being based on protein dysfunction. As many of the proteins associated with the conformational diseases are also involved in recognition, regulation, and cell signaling, it has been hypothesized that many of them are IDPs. In other words, according to this the "disorder in disorders" or D² concept, IDPs are abundantly involved in the development of the conformational diseases, which therefore may originate from the misidentification, misregulation, and missignaling due to the misfolding of causative IDPs [39].

To support this hypothesis, three approaches were elaborated for estimating the abundance of IDPs in various pathological conditions. The first approach is based on the assembly of specific datasets of proteins associated with a given disease and the computational analysis of these datasets using a number of disorder predictors [25, 39, 95, 96]. In essence, this is an analysis of individual proteins extended to a set of independent proteins. A second approach utilized network of genetic diseases where the related proteins are interlinked within one disease and between different diseases [41]. A third approach is based on the evaluation of the association between a particular protein function (including the disease-specific functional keywords) with the level of intrinsic disorder in a set of proteins known to carry out this function [13–15]. These three approaches are briefly presented below.

The easiest way to evaluate the abundance of intrinsic disorder in a given disease is based on a simple two-stage protocol, where a set of disease-related proteins is first assembled by searching various databases and then the collected group of proteins is analyzed for intrinsic disorder. The depth of this analysis is based on the breadth of the

search for the disease-related proteins and on the number of different computational tools utilized to find disordered proteins/regions [25, 39, 44, 95–97]. Using this approach, it has been shown that many proteins associated with cancer, neurodegenerative diseases, and cardiovascular disease are highly disordered, being depleted in major order-promoting residues (Trp, Phe, Tyr, Ile, and Val) and enriched in some disorder-promoting residues (Arg, Gln, Ser, Pro, and Glu). High level of intrinsic disorder and a substantial number of potential interaction sites were also found using a set of computational tools. Many proteins were predicted to be wholly disordered. Overall, these studies clearly showed that intrinsic disorder is highly prevalent in proteins associated with human diseases, being comparable with that of signaling proteins and significantly exceeding the levels of intrinsic disorder in eukaryotic and in nonhomologous, structured proteins.

Unfoldome of human genetic diseases was assembled via the analysis of a specific network which was built to estimate whether human genetic diseases and the corresponding disease genes are related to each other at a higher level of cellular and organism organization. This network represented a bipartite graph with a network of genetic diseases, the human disease network (HDN), where two diseases were directly linked if there was a gene that was directly related to both of them, and a network of disease genes, the disease gene network (DGN), where two genes were directly linked if there was a disease to which they were both directly related [98]. This framework, called the human diseasome, systematically linked the human disease phenome (which includes all the human genetic diseases) with the human disease genome (which contains all the disease-related genes) [98]. The analysis of HDN revealed that of 1284 genetic diseases, 867 had at least one link to other diseases, and 516 diseases formed a giant component, suggesting that the genetic origins of most diseases, to some extent, were shared with other diseases. In the DGN, 1377 of 1777 disease genes were shown to be connected to other disease genes, and 903 genes belonged to a giant cluster HDN. The vast majority of genes associated with genetic diseases was nonessential and showed no tendency to encode hub proteins (i.e., proteins having multiple interactions) [98]. The large-scale analysis of the abundance of intrinsic disorder in transcripts of the various disease-related genes was performed using a set of computational tools which uncovers several important features [41, 44]: (a) intrinsic disorder is common in proteins associated with many human genetic diseases; (b) different disease classes vary in the IDP contents of their associated proteins; (c) molecular recognition features, which are relatively short loosely structured protein regions within mostly disordered sequences and which gain structure upon binding to partners, are common in the diseasome, and their abundance correlates with the intrinsic disorder level; (d) some disease classes have a significant fraction of genes affected by alternative splicing, and the alternatively spliced regions in the corresponding proteins are predicted to be highly disordered and in some diseases contain a significant number of molecular recognition features, MoRFs; (e) correlations were found

among the various diseasome graph-related properties and intrinsic disorder. In agreement with earlier studies, hub proteins were shown to be more disordered.

Another approach is a computational tool elaborated for the evaluation of a correlation between the functional annotations in the SWISSPROT database and the predicted intrinsic disorder was elaborated [13–15]. The approach is based on the hypothesis that if a function described by a given keyword relies on intrinsic disorder, then the keyword-associated protein would be expected to have a greater level of predicted disorder compared to the protein randomly chosen from the SWISSPROT. To test this hypothesis, functional keywords associated with 20 or more proteins in SWISSPROT were found and corresponding keyword-associated datasets of proteins were assembled. Next, for each such a keyword-associated set, a length-matching set of random proteins was drawn from the SWISSPROT, and order-disorder predictions were carried out for the keyword-associated sets and for the random sets [13–15]. The application of this tool revealed that out of 710 SWISSPROT keywords, 310 functional keywords were associated with ordered proteins, 238 functional keywords were attributed to disordered proteins, and the remainder 162 keywords yield ambiguity in the likely function-structure associations [13–15]. It has been also shown that keywords describing various diseases were strongly correlated with proteins predicted to be disordered. Contrary to this, no disease-associated proteins were found to be strongly correlated with absence of disorder [14].

3. Experimental Tools for the Unfoldome Discovery

3.1. Enrichment of Cell Extracts in Extended IPDs by Acid Treatment. Extended IDPs, being characterized by high percentages of charged residues, do not undergo large-scale structural changes at low pH [55]. As a result, many of these proteins were shown to remain soluble under these extreme conditions [99, 100]. On the contrary, the protonation of negatively charged side chains in ordered proteins is commonly accompanied by protein denaturation or unfolding [101–104]. Unlike IDPs, the acidic pH-induced denatured conformations of structured proteins contain larger number of hydrophobic residues. The pH-induced exposure of these normally buried hydrophobic residues makes "A" states (i.e., partially folded conformations induced by acidic pH) of globular proteins "sticky," leading to their aggregation and precipitation.

Analysis or literature data revealed a set of 29 proteins which do not precipitate during perchloric acid (PCA) or trichloroacetic acid (TCA) treatment of cell extracts and this resistance to PCA or TCA treatment was utilized for their isolation [40]. However, 14 of these PCA/TCA-soluble proteins were experimentally determined to be totally unstructured, 6 were structured, and 9 had not been structurally characterized, suggesting that at least 50% of the proteins isolated by virtue of their resistance to PCA or TCA could be expected to be totally unstructured [40]. To gain more information on

the abundance of intrinsic disorder in acid-soluble proteins, their sequences were analyzed using two binary predictors of intrinsic disorder, CH-plot [45] and CDF analysis [30], both of which perform binary classification of whole proteins as either mostly disordered or mostly ordered, where mostly ordered indicates proteins that contain more ordered residues than disordered residues and mostly disordered indicates proteins that contain more disordered residues than ordered residues. The results of this analysis revealed an excellent correlation between experiment and prediction: the majority of proteins experimentally shown to be structured or unfolded were predicted to be ordered or intrinsically disordered, respectively, by both predictors. Additionally, three of four experimentally uncharacterized proteins were predicted to be wholly disordered by both classifiers. Thus, a combination of experimental and computational approaches suggested that ~70% of acid soluble proteins isolated based on their resistance to PCA or TCA could be expected to be totally unstructured [40].

Based on these observations it was suggested that indifference to acid treatment represents one of the characteristic properties of extended IDPs, which can result in the substantial enrichment of IDPs in the soluble fraction after the acid treatment, and, therefore, can be exploited to develop standard protocols for isolating and studying IDPs on a proteomic scale [40]. In agreement with this hypothesis, treatment of *E.coli* cell extracts with 1% PCA resulted in a total protein reduction of ~30 000-fold when compared to the total soluble extract, and 3% PCA was sufficient to denature and precipitate all nonresistant proteins because higher PCA concentrations did not result in further yield reductions [40]. Treatment with 3% TCA resulted in a yield similar to 1% PCA. The acid-soluble fractions from the *E.coli* extracts were visualized using 2D SDS-PAGE, which revealed that a substantial number of *E. coli* proteins were resistant to acid denaturation and concomitant precipitation. In fact, this analysis revealed that 158 proteins remained soluble in the presence of 5% PCA [40]. This suggests that ~110 of the PCA-soluble proteins could be expected to be totally unstructured (based on the assumption that ~ 70% acid-stable proteins are totally unstructured, see above). This number compares favorably with the 85 to 196 totally disordered proteins estimated to be present in the *E. coli* proteome [30]. Therefore, treating total protein extracts with 3–5% PCA or TCA and determining the identities of the soluble proteins could form the basis for uncovering unfoldomes in various organisms.

3.2. Enrichment of Cell Extracts in Extended IPDs by Heat Treatment. Several IDPs were shown to possess high resistance toward heat denaturation and aggregation. In fact, the solubility and limited secondary structure of extended IDPs, such as p21, p27, α-synuclein, prothymosin α, and phosphodiesterase γ subunit, were virtually unaltered by heating to 90°C [31, 32, 63, 100, 105–111]. This resistance to thermal aggregation, which likely originates from the low mean hydrophobicity and high net charge characteristic of extended IDPs, has been utilized for the purification of these

proteins [63, 112–115]. The indifference to heat treatment was proposed as an analytical tool for the evaluation of the abundance of extended IDPs in various proteomes [30, 57].

Recently, the extracts of NIH3T3 mouse fibroblasts were heated at a variety of temperatures and analyzed by SDS−PAGE to determine the extent of protein precipitation under these conditions [57]. In agreement with previous studies, this analysis revealed that the increase in the incubation temperature was accompanied by the decrease in the amount of soluble proteins. In fact, 375, 388, and 198 proteins (287, 304, and 124 nonredundant proteins) were identified by MALDI-TOF/TOF mass spectrometry from 584, 472, and 269 spots on 2D gels obtained for cell extracts treated at 4, 60, and 98°C, respectively. These nonredundant proteins were further analyzed using a set of bioinformatics tools. In this study, proteins were classified as IDPs (proteins having an average PONDR VLXT score >0.5 and proteins having an average PONDR score of 0.32–0.5 and possessing a high-mean net charge and low-mean hydrophobicity), intrinsically folded proteins (IFPs, proteins having an average PONDR score <0.32), or mixed ordered/disordered proteins (MPs, proteins that did not meet the described above criteria for IUPs or IFPs) [57]. The analysis clearly showed that heat treatment resulted in an enrichment of IDPs and depletion of MPs and IFPs. In fact, although IDPs comprised only 11.8% of the proteins identified in the untreated cell extract (4°C), their relative population increased to 41.9% after the heat treatment at 98°C. On the other hand, MPs and IFPs, which comprised 42.8 and 45.4% of proteins in the untreated cell extract, were substantially depleted to 27.4 and 30.6%, respectively, after heat treatment at 98°C [57].

3.3. Finding IDPs by the Combination of Native and 8 M Urea Electrophoresis of Heat-Treated Proteins. The fact that extended IDPs are characterized by heat stability and structural indifference to chemical denaturation was recently utilized in novel 2-D gel-electrophoresis technique which consists of the combination of native and 8 M urea electrophoresis of heat-treated proteins [56]. The rationales for this approach are considered below. As discussed above, extended IDPs are often heat-stable as demonstrated for Csd1 [112], MAP2 [114], α-synuclein [63, 107], its familial Parkinson's disease-related mutants [108], β- and γ-synucleins [110], stathmin [113], p21$^{\text{Cip 1}}$ [31], prothymosin α [100], C-terminal domain of caldesmon [111], and phosphodiesterase γ [109]. Therefore, heat treatment should lead to a decent initial separation of the extended IDPs from globular proteins, the vast majority of which are known to aggregate and precipitate at high temperatures. In the native gel, IDPs and rare heat-stable globular proteins will then be separated according to their charge/mass ratios. Since the extended IDPs are as unfolded in 8 M urea as under native conditions, they are expected to run the same distance in the second dimension and end up along the diagonal. Heat-stable globular proteins will unfold in urea, slow down in the second direction due to the increased size, and therefore will accumulate above the diagonal (see Figure 4 for the schematic representation of this technique). This difference
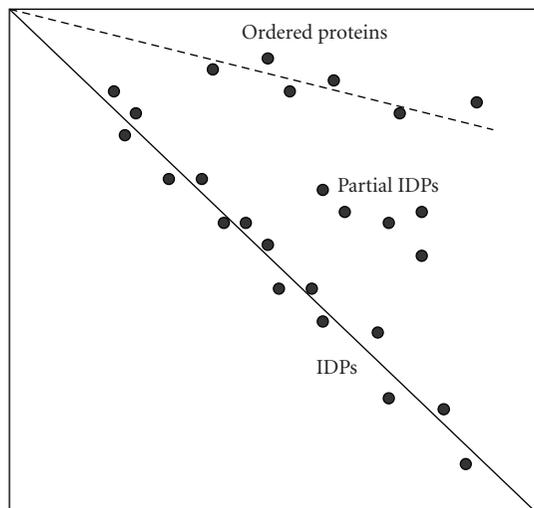
FIGURE 4: Schematic representation of the native/8 M urea 2D electrophoresis for separation of extended IDPs and globular proteins. A *continuous line* marks the diagonal of the gel to where IDPs run. A *dashed line* marks the position of globular proteins. A few proteins with a mixture of ordered and disordered regions are also indicated as "Partial IDPs".

in conformational behavior between ordered proteins and IDPs can lead to their effective separation, enabling the IDP identification by mass-spectrometry [56].

The usefulness of this approach has been validated via the analysis of a set of 10 experimentally characterized IDPs (stathmin, MAP2c, Mypt1-(304–511), ERD10, α-casein, β-casein, α-synuclein, CSD1, Bob-1, and DARPP32) and a set of 4 globular control proteins (fetuin, IPMDH, BSA, ovalbumin) [56]. The analysis revealed that IDPs ran at, or very near, the diagonal of the second (denaturing) gel, whereas globular proteins remained way above the diagonal, clearly showing that the proposed 2D electrophoresis is able to separate IDPs and globular proteins as predicted [56].

Next, heat-treated extracts of *E. coli* and *S. cerevisiae* were analyzed by this 2D electrophoresis [56]. This analysis revealed that more *S. cerevisiae* proteins were seen in the diagonal in agreement with predictions that the frequency of protein disorder increases with increasing complexity of the organisms [5, 28–30]. At the next step, some spots at and above the diagonal were identified by mass-spectrometry and the intrinsic disorder propensity of the identified proteins was estimated by PONDR VLXT [56]. This analysis revealed that the amount of predicted disorder in proteins located at the diagonal positions was very high ($52.1 \pm 14.1\%$), noticeably exceeding that of typical IDPs such as α-synuclein (37.1%) and α-casein (41.15%) [56]. Although many of the "diagonal" proteins have never been structurally character-ized, literature data were available for some of them. The list of such previously characterized IDPs identified in this study includes ribosomal proteins, GroES, and acyl carrier protein. The majority of proteins above the diagonal were found to be enzymes (e.g., superoxide dismutase), which are known to require a well-defined structure for function

[56]. Based on these finding it has been concluded that the proposed 2D electrophoresis is suitable for the proteome-wide identification of IDPs.

4. Concluding Remarks

Intrinsic disorder is highly abundant in nature. According to the genome-based bioinformatics predictions, significant fraction of any given proteome belongs to the class of IDPs. These proteins possess numerous vital functions. Many proteins associated with various human diseases are intrin-sically disordered too. High degree of association between protein intrinsic disorder and maladies is due to structural and functional peculiarities of IDPs and IDRs, which are typically involved in cellular regulation, recognition, and signal transduction. As the number of IDPs is very large and as many of these proteins are interlinked, the concepts of the unfoldome and unfoldomics were introduced. IDPs, especially their extended forms, are characterized by several unique features that can be used for isolation of these proteins from the cell extracts. The corresponding proteomic techniques utilize specific high resistance of IDPs against extreme pH and high temperature, as well as their structural indifference to chemical denaturation. At the computational side, several specific features of the IDP amino acid sequences provide a solid background for the reliable identification of these proteins at the proteome level. These proteomic-scale identification and characterization of IDPs are needed to advance our knowledge in this important field.

Acknowledgments

References

[1] E. Fischer, "Einfluss der configuration auf die wirkung der enzyme," *Berichte der Deutschen Chemischen Gesellschaft*, vol. 27, no. 3, pp. 2985–2993, 1894.

[2] R. U. Lemieux and U. Spohr, "How Emil Fischer was led to the lock and key concept for enzyme specificity," *Advances in Carbohydrate Chemistry and Biochemistry*, vol. 50, pp. 1–20, 1994.

[3] C. M. Dobson, "Protein misfolding, evolution and disease," *Trends in Biochemical Sciences*, vol. 24, no. 9, pp. 329–332, 1999.

[4] A. K. Dunker, C. J. Oldfield, J. Meng, et al., "The unfoldomics decade: an update on intrinsically disordered proteins," *BMC Genomics*, vol. 9, supplement 2, article S1, 2008.

[5] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradović, "Intrinsic disorder and protein function," *Biochemistry*, vol. 41, no. 21, pp. 6573–6582, 2002.

[6] A. K. Dunker, C. J. Brown, and Z. Obradović, "Identification and functions of usefully disordered proteins," *Advances in Protein Chemistry*, vol. 62, pp. 25–49, 2002.

[7] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky, "Flexible nets: the roles of intrinsic disorder in protein interaction networks," *FEBS Journal*, vol. 272, no. 20, pp. 5129–5148, 2005.

[8] A. K. Dunker, J. D. Lawson, C. J. Brown, et al., "Intrinsically disordered protein," *Journal of Molecular Graphics and Modelling*, vol. 19, no. 1, pp. 26–59, 2001.

[9] A. K. Dunker and Z. Obradović, "The protein trinity—linking function and disorder," *Nature Biotechnology*, vol. 19, no. 9, pp. 805–806, 2001.

[10] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradović, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder and functional proteomics," *Biophysical Journal*, vol. 92, no. 5, pp. 1439–1456, 2007.

[11] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, "Function and structure of inherently disordered proteins," *Current Opinion in Structural Biology*, vol. 18, no. 6, pp. 756–764, 2008.

[12] A. K. Dunker and V. N. Uversky, "Signal transduction via unstructured protein conduits," *Nature Chemical Biology*, vol. 4, no. 4, pp. 229–230, 2008.

[13] S. Vucetic, H. Xie, L. M. Iakoucheva, et al., "Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1899–1916, 2007.

[14] H. Xie, S. Vucetic, L. M. Iakoucheva, et al., "Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1917–1932, 2007.

[15] H. Xie, S. Vucetic, L. M. Iakoucheva, et al., "Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1882–1898, 2007.

[16] M. S. Cortese, V. N. Uversky, and A. Keith Dunker, "Intrinsic disorder in scaffold proteins: getting more from less," *Progress in Biophysics and Molecular Biology*, vol. 98, no. 1, pp. 85–106, 2008.

[17] R. B. Russell and T. J. Gibson, "A careful disorderliness in the proteome: sites for interaction and targets for future therapies," *FEBS Letters*, vol. 582, no. 8, pp. 1271–1275, 2008.

[18] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Qu, V. N. Uversky, and A. K. Dunker, "Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners," *BMC Genomics*, vol. 9, supplement 1, article S1, 2008.

[19] P. Tompa and P. Csermely, "The role of structural disorder in the function of RNA and protein chaperones," *FASEB Journal*, vol. 18, no. 11, pp. 1169–1175, 2004.

[20] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling," *Journal of Molecular Recognition*, vol. 18, no. 5, pp. 343–384, 2005.

[21] P. E. Wright and H. J. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal of Molecular Biology*, vol. 293, no. 2, pp. 321–331, 1999.

[22] P. Tompa, "Intrinsically unstructured proteins," *Trends in Biochemical Sciences*, vol. 27, no. 10, pp. 527–533, 2002.

[23] V. N. Uversky, "Natively unfolded proteins: a point where biology waits for physics," *Protein Science*, vol. 11, no. 4, pp. 739–756, 2002.

[24] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 3, pp. 197–208, 2005.

[25] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradović, and A. K. Dunker, "Intrinsic disorder in cell-signaling and cancer-associated proteins," *Journal of Molecular Biology*, vol. 323, no. 3, pp. 573–584, 2002.

[26] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, "Function and structure of inherently disordered proteins," *Current Opinion in Structural Biology*, vol. 18, no. 6, pp. 756–764, 2008.

[27] P. Tompa, "The interplay between structure and function in intrinsically unstructured proteins," *FEBS Letters*, vol. 579, no. 15, pp. 3346–3354, 2005.

[28] A. K. Dunker, Z. Obradović, P. Romero, E. C. Garner, and C. J. Brown, "Intrinsic protein disorder in complete genomes," *Genome Informatics*, vol. 11, pp. 161–171, 2000.

[29] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life," *Journal of Molecular Biology*, vol. 337, no. 3, pp. 635–645, 2004.

[30] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, and A. K. Bunker, "Comparing and combining predictors of mostly disordered proteins," *Biochemistry*, vol. 44, no. 6, pp. 1989–2000, 2005.

[31] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright, "Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 21, pp. 11504–11509, 1996.

[32] E. R. Lacy, I. Filippov, W. S. Lewis, et al., "p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding," *Nature Structural and Molecular Biology*, vol. 11, no. 4, pp. 358–364, 2004.

[33] E. R. Lacy, Y. Wang, J. Post, et al., "Molecular basis for the specificity of p27 toward cyclin-dependent kinases that regulate cell division," *Journal of Molecular Biology*, vol. 349, no. 4, pp. 764–773, 2005.

[34] H. J. Dyson and P. E. Wright, "Coupling of folding and binding for unstructured proteins," *Current Opinion in Structural Biology*, vol. 12, no. 1, pp. 54–60, 2002.

[35] C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, and A. K. Dunker, "Coupled folding and binding with alpha-helix-forming molecular recognition elements," *Biochemistry*. In press.

[36] Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky, and A. K. Dunker, "Mining α-helix-forming molecular recognition features with cross species sequence alignments," *Biochemistry*, vol. 46, no. 47, pp. 13468–13477, 2007.

[37] A. Mohan, *MoRFs: A Dataset of Molecular Recognition Features*, The School of Informatics, Indiana University, Indianapolis, Ind, USA, 2006.

[38] V. Vacic, C. J. Oldfield, A. Mohan, et al., "Characterization of molecular recognition features, MoRFs, and their binding partners," *Journal of Proteome Research*, vol. 6, no. 6, pp. 2351–2366, 2007.

[39] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Intrinsically disordered proteins in human diseases: introducing the D2 concept," *Annual Review of Biophysics*, vol. 37, pp. 215–246, 2008.

[40] M. S. Cortese, J. P. Baird, V. N. Uversky, and A. K. Dunker, "Uncovering the unfoldome: enriching cell extracts for unstructured proteins by acid treatment," *Journal of Proteome Research*, vol. 4, no. 5, pp. 1610–1618, 2005.

[41] U. Midic, C. J. Oldfield, A. K. Keith, Z. Obradović, and V. N. Uversky, "Protein disorder in the human diseasome: unfoldomics of human genetic diseases," *BMC Genomics*, vol. 10, supplement 1, article S12, 2009.

[42] H. Winkler, V.u.U.d.P.i.P.-u.T.J.V.F., 1920.

[43] J. Lederberg and A. T. McCray, "'Ome' sweet 'omics'—a genealogical treasury of words," *The Scientist*, vol. 15, no. 7, article 8, 2001.

[44] V. N. Uversky, C. J. Oldfield, U. Midic, et al., "Unfoldomics of human diseases: linking protein intrinsic disorder with diseases," *BMC Genomics*, vol. 10, supplement 1, article S7, 2009.

[45] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?" *Proteins*, vol. 41, no. 3, pp. 415–427, 2000.

[46] F. Ferron, S. Longhi, B. Canard, and D. Karlin, "A practical overview of protein disorder prediction methods," *Proteins*, vol. 65, no. 1, pp. 1–14, 2006.

[47] Z. Dosztanyi, M. Sandor, P. Tompa, and I. Simon, "Prediction of protein disorder at the domain level," *Current Protein and Peptide Science*, vol. 8, no. 2, pp. 161–171, 2007.

[48] Z. Dosztanyi and P. Tompa, "Prediction of protein disorder," *Methods in Molecular Biology*, vol. 426, pp. 103–115, 2008.

[49] S. J. Hubbard, R. J. Beynon, and J. M. Thornton, "Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures," *Protein Engineering*, vol. 11, no. 5, pp. 349–359, 1998.

[50] L. M. Iakoucheva, A. L. Kimzey, C. D. Masselon, R. D. Smith, A. K. Dunker, and E. J. Ackerman, "Aberrant mobility phenomena of the DNA repair protein XPA," *Protein Science*, vol. 10, no. 7, pp. 1353–1362, 2001.

[51] R. Reeves and M. S. Nissen, "Purification and assays for high mobility group HMG-I(Y) protein function," *Methods in Enzymology*, vol. 304, pp. 155–188, 1999.

[52] G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese, and A. K. Dunker, "Natively disordered proteins," in *Handbook of Protein Folding*, J. Buchner and T. Kiefhaber, Eds., pp. 271–353, Wiley-VCH, Weinheim, Germany, 2005.

[53] V. Receveur-Bréchot, J.-M. Bourhis, V. N. Uversky, B. Canard, and S. Longhi, "Assessing protein disorder and induced folding," *Proteins*, vol. 62, no. 1, pp. 24–45, 2006.

[54] V. N. Uversky, "What does it mean to be natively unfolded?" *European Journal of Biochemistry*, vol. 269, no. 1, pp. 2–12, 2002.

[55] V. N. Uversky, "Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, ph, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding," *Protein Journal*, vol. 28, no. 7-8, pp. 305–325, 2009.

[56] V. Csizmók, E. Szollosi, P. Friedrich, and P. Tompa, "A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins," *Molecular and Cellular Proteomics*, vol. 5, no. 2, pp. 265–273, 2006.

[57] C. A. Galea, V. R. Pagala, J. C. Obenauer, C.-G. Park, C. A. Slaughter, and R. W. Kriwacki, "Proteomic studies of the intrinsically unstructured mammalian proteome," *Journal of Proteome Research*, vol. 5, no. 10, pp. 2839–2848, 2006.

[58] P. Romero, Z. Obradović, C. Kissinger, J. E. Villafranca, and A. K. Dunker, "Identifying disordered regions in proteins from amino acid sequence," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 1, pp. 90–95, 1997.

[59] P. Romero, Z. Obradović, and K. Dunker, "Sequence data analysis for long disordered regions prediction in the calcineurin family," *Genome Informatics*, vol. 8, pp. 110–124, 1997.

[60] Q. Xie, G. E. Arnold, P. Romero, Z. Obradović, E. Garner, and A. K. Dunker, "The sequence attribute method for determining relationships between sequence and protein disorder," *Genome Informatics*, vol. 9, pp. 193–200, 1998.

[61] P. Romero, Z. Obradović, C. R. Kissinger, et al., "Thousands of proteins likely to have long disordered regions," *Pacific Symposium on Biocomputing*, pp. 437–448, 1998.

[62] K. Gast, H. Damaschun, K. Eckert, et al., "Prothymosin α: a biologically active protein with random coil conformation," *Biochemistry*, vol. 34, no. 40, pp. 13211–13218, 1995.

[63] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury Jr., "NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded," *Biochemistry*, vol. 35, no. 43, pp. 13709–13715, 1996.

[64] H. C. Hemmings Jr., A. C. Nairn, D. W. Aswad, and P. Greengard, "DARPP-32, a dopamine- and adenosine 3′:5′-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus," *Journal of Neuroscience*, vol. 4, no. 1, pp. 99–110, 1984.

[65] R. J. Williams, "The conformational mobility of proteins and its functional significance," *Biochemical Society Transactions*, vol. 6, no. 6, pp. 1123–1126, 1978.

[66] A. K. Dunker, E. Garner, S. Guilliot, et al., "Protein disorder and the evolution of molecular recognition: theory, predictions and observations," *Pacific Symposium on Biocomputing*, pp. 473–484, 1998.

[67] E. Garner, P. Cannon, P. Romero, Z. Obradović, and A. K. Dunker, "Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization," *Genome Informatics*, vol. 9, pp. 201–213, 1998.

[68] M. Vihinen, E. Torkkila, and P. Riikonen, "Accuracy of protein flexibility predictions," *Proteins*, vol. 19, no. 2, pp. 141–149, 1994.

[69] R. M. Williams, Z. Obradović, V. Mathura, et al., "The protein non-folding problem: amino acid determinants of intrinsic order and disorder," *Pacific Symposium on Biocomputing*, pp. 89–100, 2001.

[70] P. Romero, Z. Obradović, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence complexity of disordered protein," *Proteins*, vol. 42, no. 1, pp. 38–48, 2001.

[71] X. Li, Z. Obradović, C. J. Brown, E. C. Garner, and A. K. Dunker, "Comparing predictors of disordered protein," *Genome Informatics*, vol. 11, pp. 172–184, 2000.

[72] P. Radivojac, Z. Obradović, D. K. Smith, et al., "Protein flexibility and intrinsic disorder," *Protein Science*, vol. 13, no. 1, pp. 71–80, 2004.

[73] S. Vucetic, C. J. Brown, A. K. Dunker, and Z. Obradović, "Flavors of protein disorder," *Proteins*, vol. 52, no. 4, pp. 573–584, 2003.

[74] X. Li, P. Romero, M. Rani, A. K. Dunker, and Z. Obradović, "Predicting protein disorder for N-, C-, and internal regions," *Genome Informatics*, vol. 10, pp. 30–40, 1999.

[75] K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, and Z. Obradović, "Optimizing long intrinsic disorder predictors with protein evolutionary information," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 1, pp. 35–60, 2005.

[76] Z. Obradović, K. Peng, S. Vucetic, P. Radivojac, and A. K. Bunker, "Exploiting heterogeneous sequence properties improves prediction of protein disorder," *Proteins*, vol. 61, supplement 7, pp. 176–182, 2005.

[77] K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker, and Z. Obradović, "Length-dependent prediction of protein in intrinsic disorder," *BMC Bioinformatics*, vol. 7, article 208, 2006.

[78] J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, et al., "FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded," *Bioinformatics*, vol. 21, no. 16, pp. 3435–3438, 2005.

[79] R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson, "GlobPlot: exploring protein sequences for globularity and disorder," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3701–3708, 2003.

[80] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell, "Protein disorder prediction: implications for structural proteomics," *Structure*, vol. 11, no. 11, pp. 1453–1459, 2003.

[81] D. T. Jones and J. J. Ward, "Prediction of disordered regions in proteins from position specific score matrices," *Proteins*, vol. 53, supplement 6, pp. 573–578, 2003.

[82] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones, "The DISOPRED server for the prediction of protein disorder," *Bioinformatics*, vol. 20, no. 13, pp. 2138–2139, 2004.

[83] K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones, "Protein structure prediction servers at University College London," *Nucleic Acids Research*, vol. 33, supplement 2, pp. W36–W38, 2005.

[84] Z. Dosztanyi, V. Csizmok, P. Tompa, and I. Simon, "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content," *Bioinformatics*, vol. 21, no. 16, pp. 3433–3434, 2005.

[85] O. V. Galzitskaya, S. O. Garbuzynskiy, and M. Y. Lobanov, "FoldUnfold: web server for the prediction of disordered regions in protein chain," *Bioinformatics*, vol. 22, no. 23, pp. 2948–2949, 2006.

[86] Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf, "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins," *Bioinformatics*, vol. 21, no. 16, pp. 3369–3376, 2005.

[87] C.-T. Su, C.-Y. Chen, and Y.-Y. Ou, "Protein disorder prediction by condensed PSSM considering propensity for order or disorder," *BMC Bioinformatics*, vol. 7, article 319, 2006.

[88] C. T. Su, C. Y. Chen, and C. M. Hsu, "iPDA: integrated protein disorder analyzer," *Nucleic Acids Research*, vol. 35, web server issue, pp. W465–W472, 2007.

[89] A. Vullo, O. Bortolamil, G. Pollastri, and S. C. E. Tosatto, "Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines," *Nucleic Acids Research*, vol. 34, web server issue, pp. W164–W168, 2006.

[90] T. Ishida and K. Kinoshita, "PrDOS: prediction of disordered protein regions from amino acid sequence," *Nucleic Acids Research*, vol. 35, web server issue, pp. W460–W464, 2007.

[91] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, and A. K. Dunker, "Predicting intrinsic disorder in proteins: an overview," *Cell Research*, vol. 19, no. 8, pp. 929–949, 2009.

[92] A. Mohan, W. J. Sullivan Jr., P. Radivojac, A. K. Dunker, and V. N. Uversky, "Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes," *Molecular BioSystems*, vol. 4, no. 4, pp. 328–340, 2008.

[93] K. I. Rantalainen, V. N. Uversky, P. Permi, N. Kalkkinen, A. K. Dunker, and K. Mäkinen, "Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core," *Virology*, vol. 377, no. 2, pp. 280–288, 2008.

[94] E. Hébrard, Y. Bessin, T. Michon, et al., "Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses," *Virology Journal*, vol. 6, article 23, 2009.

[95] V. N. Uversky, A. Roman, C. J. Oldfield, and A. K. Dunker, "Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs," *Journal of Proteome Research*, vol. 5, no. 8, pp. 1829–1842, 2006.

[96] Y. Cheng, T. LeGall, C. J. Oldfield, A. K. Dunker, and V. N. Uversky, "Abundance of intrinsic disorder in protein associated with cardiovascular disease," *Biochemistry*, vol. 45, no. 35, pp. 10448–10460, 2006.

[97] V. N. Uversky, "Intrinsic disorder in proteins associated with neurodegenerative diseases," *Frontiers in Bioscience*, vol. 14, pp. 5188–5238, 2009.

[98] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 21, pp. 8685–8690, 2007.

[99] V. N. Uversky, "A protein-chameleon: conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders," *Journal of Biomolecular Structure and Dynamics*, vol. 21, no. 2, pp. 211–234, 2003.

[100] V. N. Uversky, J. R. Gillespie, I. S. Millett, et al., "Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic pH," *Biochemistry*, vol. 38, no. 45, pp. 15009–15016, 1999.

[101] Y. Goto, L. J. Calciano, and A. L. Fink, "Acid-induced folding of proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 2, pp. 573–577, 1990.

[102] Y. Goto and A. L. Fink, "Phase diagram for acidic conformational states of apomyoglobin," *Journal of Molecular Biology*, vol. 214, no. 4, pp. 803–805, 1990.

[103] Y. Goto, N. Takahashi, and A. L. Fink, "Mechanism of acid-induced folding of proteins," *Biochemistry*, vol. 29, no. 14, pp. 3480–3488, 1990.

[104] A. L. Fink, L. J. Calciano, Y. Goto, T. Kurotsu, and D. R. Palleros, "Classification of acid denaturation of proteins: intermediates and unfolded states," *Biochemistry*, vol. 33, no. 41, pp. 12504–12511, 1994.

[105] E. A. Bienkiewicz, J. N. Adkins, and K. J. Lumb, "Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27Kip1," *Biochemistry*, vol. 41, no. 3, pp. 752–759, 2002.

[106] L. Hengst, V. Dulic, J. M. Slingerland, E. Lees, and S. I. Reed, "A cell cycle-regulated inhibitor of cyclin-dependent kinases,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 91, no. 12, pp. 5291–5295, 1994.

[107] V. N. Uversky, J. Li, and A. L. Fink, "Evidence for a partially folded intermediate in α-synuclein fibril formation," *Journal of Biological Chemistry*, vol. 276, no. 14, pp. 10737–10744, 2001.

[108] J. Li, V. N. Uversky, and A. L. Fink, "Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human α-synuclein," *Biochemistry*, vol. 40, no. 38, pp. 11604–11613, 2001.

[109] V. N. Uversky, S. E. Permyakov, V. E. Zagranichny, et al., "Effect of zinc and temperature on the conformation of the γ subunit of retinal phosphodiesterase: a natively unfolded protein," *Journal of Proteome Research*, vol. 1, no. 2, pp. 149–159, 2002.

[110] V. N. Uversky, J. Li, P. Souillac, et al., "Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of α-synuclein assembly by β- and γ-synucleins," *Journal of Biological Chemistry*, vol. 277, no. 14, pp. 11970–11978, 2002.

[111] S. E. Permyakov, I. S. Millett, S. Doniach, E. A. Permyakov, and V. N. Uversky, "Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin," *Proteins*, vol. 53, no. 4, pp. 855–862, 2003.

[112] M. Häckel, T. Konno, and H.-J. Hinz, "A new alternative method to quantify residual structure in 'unfolded' proteins," *Biochimica et Biophysica Acta*, vol. 1479, no. 1-2, pp. 155–165, 2000.

[113] L. D. Belmont and T. J. Mitchison, "Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules," *Cell*, vol. 84, no. 4, pp. 623–631, 1996.

[114] M. A. Hernandez, J. Avila, and J. M. Andreu, "Physico-chemical characterization of the heat-stable microtubule-associated protein MAP2," *European Journal of Biochemistry*, vol. 154, no. 1, pp. 41–48, 1986.

[115] C. Kalthoff, "A novel strategy for the purification of recombinantly expressed unstructured protein domains," *Journal of Chromatography B*, vol. 786, no. 1-2, pp. 247–254, 2003.

Review Article

Proteomics of Plant Pathogenic Fungi

Raquel González-Fernández,[1] Elena Prats,[2] and Jesús V. Jorrín-Novo[1]

[1] *Agricultural and Plant Biochemistry and Proteomics Research Group, Department of Biochemistry and Molecular Biology,*
 University of Córdoba, 14071 Córdoba, Spain
[2] *CSIC, Institute of Sustainable Agriculture, 14080 Córdoba, Spain*

Correspondence should be addressed to Raquel González-Fernández, q42gofer@uco.es

Plant pathogenic fungi cause important yield losses in crops. In order to develop efficient and environmental friendly crop protection strategies, molecular studies of the fungal biological cycle, virulence factors, and interaction with its host are necessary. For that reason, several approaches have been performed using both classical genetic, cell biology, and biochemistry and the modern, holistic, and high-throughput, omic techniques. This work briefly overviews the tools available for studying Plant Pathogenic Fungi and is amply focused on MS-based Proteomics analysis, based on original papers published up to December 2009. At a methodological level, different steps in a proteomic workflow experiment are discussed. Separate sections are devoted to fungal descriptive (intracellular, subcellular, extracellular) and differential expression proteomics and interactomics. From the work published we can conclude that Proteomics, in combination with other techniques, constitutes a powerful tool for providing important information about pathogenicity and virulence factors, thus opening up new possibilities for crop disease diagnosis and crop protection.

1. Introduction: Plant Parasitic Fungi

Fungi form a large and heterogeneous eukaryotic group of living organisms characterized by their lack of photosynthetic pigment and their chitinous cell wall. It has been estimated that the fungal kingdom contains more than 1.5 million species, but only around 100,000 have so far been described, with yeast, mold, and mushroom being the most familiar [1]. Although the majority of fungal species are saprophytes, a number of them are parasitics, in order to complete their biological cycle, animals or plants, with around 15,000 of them causing disease in plants, the majority belonging to the Ascomycetes and Basidiomycetes [2] (Table 1). Within a fungal plant pathogen species, for example, in *Fusarium oxysporum*, up to 120 different formae specialis can be found based on specificity to host species belonging to a wide range of plant families [3].

According to the type of parasitism and infection strategy, fungi are classified as biotrophic (e.g., *Blumeria graminis*), necrotophic (e.g., *Botrytis cinerea*), or hemibiotrophic (e.g., *Colletotrichum destructivum*). While the former derives nutrients from dead cells, the latter takes nutrients from the plant but does not kill it [4]. Hemibiotrophes sequentially deploy a biotrophic and then a necrotrophic mode of nutrition. Necrotrophic species tend to attack a broad range of plant species; on the contrary, biotrophes usually exhibit a high degree of specialization for individual plant species. Most biotrophic fungi are obligatory parasites, surviving only limited saprophytic phases. Differently from necrotrophes, the cultivation of biotrophic fungi succeeds only in a few exceptions, for example, *Podosphaera fusuca* [5] or *B. graminis* (M. M. Corbitt, personal communication, adapted from [5]).

Fungal diseases are, in nature, more the exception than the rule. Thus, only a limited number of fungal species are able to penetrate and invade host tissues, avoiding recognition and plant defence responses, in order to obtain nutrients from them, causing disease and sometimes host death. In agriculture, annual crop losses due to pre- and post harvest fungal diseases exceed 200 billion euros, and, in the United Stated alone, over $600 million are annually spent on fungicides [6].

TABLE 1: Main plant pathogenic fungi causing disease in plants.

Phylum	Genus	Anamorphic stage	Hosts	Disease	Example
Chytridiomycota	Olpidium		cabbage	root diseases	*O. brassicae*
	Physoderma		corn	brown spot	*P. maydis*
			alfalfa	crown wart	*P. (= Urophlyctis) alfalfae*
	Synchytrium		potato	potato wart	*S. endobioticum*
Zygomycota	Rhizopus		fruits and vegetables	bread molds and soft rot	*R. oligosporus*
	Choanephora		squash	soft rot	*C. cucurbitarum*
	Mucor		fruits and vegetables	bread mold and storage rots	*M. indicus*
Ascomycota	Taphrina		peach plum oak	leaf curl leaf blister and so forth	*T. deformans*
	Galactomyces		citrus	sour rot	*G. candidum*
	Blumeria		cereals and grasses	powdery mildew	*B. graminis*[1]
	Erysiphe	Oidium	many herbaceous plants	powdery mildew	*E. pisi*
	Leveillula		tomato	powdery mildew	*L. taurica*
	Microsphaera		lilac	powdery mildew	*M. penicullata*
	Oidium		tomato	powdery mildew	*O. neolycopersici*
	Podosphaera		apple	powdery mildew	*P. leucotricha*
	Sphaerotheca		roses and peach	powdery mildew	*S. pannosa*
	Uncinula		grape	powdery mildew	*U. necator*
	Nectria		trees	twig and stem cankers	*N. galligena*
	Gibberella		corn and small grains	foot or stalk rot	*F. graminearum*[1]
		Fusarium	several plants	vascular wilts root rots stem rots seed infections	F. oxysporum[1]
	Claviceps		grain crops	ergot	*C. purpurea*
	Ceratocystis	Chalara	oak	oak wilt	*C. fagacearum*
			stone fruit and sweet potato	cankers and root rot	*C. fimbriata*
			pineapple	butt rot	*C. paradoxa*
	Monosporascus		cucurbits	root rot and collapse	*M.* cannonballus
	Glomerella		apple	anthracnoses and bitter rot	*G. cingulata*
		Colletotrichum	many plants	anthracnoses	*C. lindemuthianum*
	Phyllachora		grasses	leaf spots	*P. graminis*
	Ophiostoma	Sporothrix and Graphium	elm	Dutch elm disease	*O. novo*-ulm
	Diaporthe			citrus melanose	*D. citri*
				eggplant fruit rot	*D. vexans*
				soybean pod and stem rot	*D. phaseolorum*
	Gaeumannomyces		grain crops and grasses	take-all disease	*G. graminis*
	Magnaporthe		rice	rice blast	*M. grisea*[1]

TABLE 1: Continued.

Phylum	Genus	Anamorphic stage	Hosts	Disease	Example
	Cryphonectria		chestnut	blight disease	*C. parasitica*
	Leucostoma		peach and other trees	canker diseases	*L. persoonii*
	Hypoxylon		poplars	canker disease	*H. mammatum*
	Rosellinia		fruit trees and vines	root diseases	*R. necatrix*
	Xylaria		trees	tree cankers and wood decay	*X. longipes*
	Eutypa		fruit trees and vines	canker	*E. armeniacae*
	Mycosphaerella	Cercospora	Banana	Sigatoka disease	*M. musicola* and *M. fijiensis*
		Septoria	cereals and grasses	leaf spots	*M. graminicola*
			strawberry	leaf spot	*M. fragariae*
	Elsinoë		citrus trees	citrus scab	*E. fawcetti*
			grape	anthracnose	*E. ampelina*
			raspberry	anthracnose	*E. veneta*
	Capnodium		most plants	sooty molds	*C. elaeophilum*
	Cochliobolus	Bipolaris	grain crops and grasses	leaf spots and root rots	*C. carbonum* and *B. maydis*
		Curvularia	grasses	leaf spots	*C. lunata*[1]
	Pyrenophora	Drechslera	cereals and grasses	leaf spots	*P. graminea*
	Setosphaera		cereals and grasses	leaf spots	*S. turcica*
	Pleospora	Stemphylium	tomato	black mold rot	*P. lycopersici* and *S. solani*
	Leptosphaeria		cabbage	black leg and foot rot	*L. maculans*[1]
	Venturia		apple	apple scab	*V. inaequalis*
			pear	pear scab	*V. pyrina*
		Cladosporium	tomato	leaf mold	*C. fulvum*
			peach and almond	scab	*C. carpophilum*
	Guignardia	Phyllosticta	grapes	black rot	*G. bidwellii*
	Apiosporina		cherries and plums	black knot	*A. morbosa*
	Hypoderma		pines	needle cast	*H. desmazierii*
	Lophodermium		pines	needle cast	*L. pinastri*
	Rhabdocline		pines	Douglas fir needle cast	*R. weirii*
	Rhytisma		maple	tar spot of leaves	*R. acerinum*
	Monilinia		stone fruit	brown rot disease	*M. fruticola*
	Sclerotinia		vegetables	white mold	*S. sclerotiorum*[1]
	Stromatinia		gladiolus	corm rot	*S. gladioli*
	Pseudopeziza		alfalfa	leaf spot	*P. trifolii*
	Diplocarpon		quince and pear	black spot	*D. maculatum*
	Talaromyces	Penicillium	fruits	blue mold rot	*P. digitatum*
		Aspergillus	seeds	bread mold and seed decays	*A. niger*[1]
	Hypocrea	Verticillium	many plants	vascular wilts	*V. dahliae*[1]
	Lewia	Alternaria	many plants	leaf spots and blights	*A. alternata*
	Setosphaera	Exserohilum	grasses	leaf spots	*E. longirostratum*
	Botryosphaeria	Sphaeropsis	apple	black rot	*S. pyriputrescens*
	Botryotinia	Botrytis	many plants	gray mold rots	*B. cinerea*[1]

Table 1: Continued.

Phylum	Genus	Anamorphic stage	Hosts	Disease	Example
	Monilinia	Monilia	stone fruits	brown rot	*M. fruticola*
	Diplocarpon	Entomosporium	pear	leaf and fruit spot	*E. mespili*
	Greeneria	Melanconium	grape	bitter rot	*M. fuligineum*
Basidiomycota	Ustilago		corn	smut	*U. maydis*[1]
			oats	loose smuts	*U. avenae*
			barley	loose smuts	*U. nuda*
			wheat	loose smuts	*U. tritici*
	Tilletia		wheat	covered smut or bunt	*T. caries*
			wheat	Karnal bunt	*T. indica*
	Urocystis		onion	smut	*U. cepulae*
	Sporisorium		sorghum	covered kernel smut	*S. sorghi*
			sorghum	loose sorghum smut	*S. cruentum*
	Sphacelotheca		sorghum	head smut	*S. reiliana*
	Cronartium		pines	blister rust	*C. ribicola*
	Gymnosporangium		apple	cedar-apple rust	*G. juniperi-virginianae*
	Hemileia		coffee	rust	*H. vastatrix*
	Melampsora		flax	rust	*M. lini*
	Phakopsora		soybeans	rust	*P. pachyrrhizi*
	Puccinia		cereals	rust	*P. recondita*
	Uromyces		beans	rust	*U. appendiculatus*[1]
	Exobasidium		ornamentals	leaf flower and stem galls	*E. japonicum*
	Athelia		many plants	Southern blight	*A. rolfsii*
		Sclerotium	onions	white rot	*S. cepivorum*
	Thanatephorus	Rhizoctonia	many plants	root and stem rots damping-off and fruit rots	*T. cucumeris and R.solani*
	Typhula		turf grasses	snow mold	*T. incarnata*
	Armillaria		trees	root rots	*A. mellea*
	Crinipellis		cacao	witches'-broom	*C. perniciosus*
	Marasmius		turf grasses	fairy ring disease	*M. oreades*
	Pleurotus		trees	white rot on logs tree stumps and living trees	*P. ostreatus*[1]
	Pholiota		trees	brown wood rot	*P. squarrosa*
	Chondrostereum		trees	silver leaf disease	*C. purpureum*
	Corticium		turf grasses	red thread disease	*C. fuciforme*
	Heterobasidion		trees	root and butt rot	*H. annosum*
	Ganoderma		trees	root and basal stem rots	*G. boninense*
	Inonotus		trees	heart rot	*I. hispidus*
	Polyporus		trees	heart rot	*P. glomeratus*
	Postia		trees	wood and root rot	*P. fragilis*

[1] These phytopathogenic fungi are named in the text.
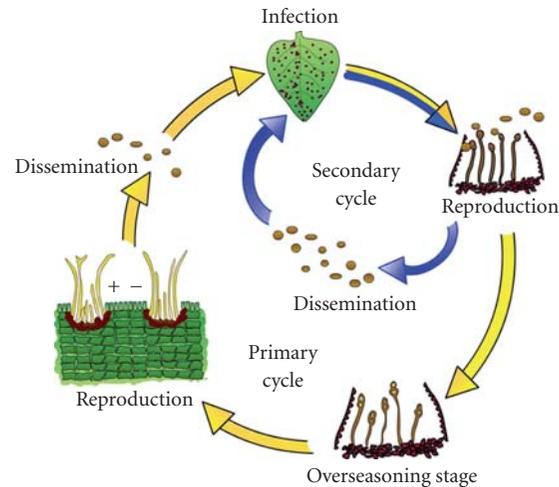
FIGURE 1: Diagram of monocyclic (yellow) and polycyclic (yellow and blue) fungi. In monocyclic diseases the fungus produces spores at the end of the season that serve as primary and only inoculum for the following year. The primary inoculum infects plants during the growth season and, at the end of the growth season, produces new spores in the infected tissues. These spores remain in the soil (overseasoning stage) and serve as the primary inoculum the following season. In polycyclic fungal pathogens, the primary inoculum often consists of the sexual (perfect) spore or, in fungi that lack the sexual stage, some other structures such as sclerotia, pseudosclerotia, or mycelium in infected tissue. This inoculum causes the primary infection and then large numbers of asexual spores (secondary inoculum) are produced at each infection site and these spores can themselves cause new (secondary) infections that produce more asexual spores for more infections.

Fungal pathogens have complicated life cycles, with both asexual and sexual reproduction, and stages involving the formation of different infective, vegetative, and reproductive structures [7]. The primary events in a disease cycle are establishment of infection, colonization (invasion), growth and reproduction, of the pathogen, dissemination of the pathogen, and survival of the pathogen in the absence of the host, that is, overwintering or oversummering (overseasoning) of the pathogen (Figure 1). However, the execution of each stage largely differs depending on the pathogen [8]. In polycyclic diseases there are several infection cycles within one, the so-called secondary cycles [3] (Figure 1).

The fungal plant interplay depends on mutual recognition, signalling, and the expression of pathogenicity and virulence factors, from the fungal side, and the existence of passive, preformed, or inducible defence mechanisms in the plant, resulting in compatible (susceptibility) or incompatible (nonhost, basal or host specific resistance) interactions. From a genetic point of view, and according to the gene-for-gene interaction hypothesis, proposed by Flor while studying flax rust [9], resistance results from the combination of a dominant avirulence (*Avr*) gene in the pathogen and a cognate resistance (*R*) gene in the host; the interaction of both gene products leads to the activation of host defence responses, such as the hypersensitive response, that arrests the growth of fungi. This hypothesis has been experimentally demonstrated for a number of pathosystems, mainly involving biotrophic fungi, with a number of avirulent genes identified [10].

A number of fungal mechanisms and molecules have been shown to contribute to fungal pathogenicity or virulence, understood as the capacity to cause damage in a host, in absolute or relative terms. Among them, cell wall degrading proteins, inhibitory proteins [11], and enzymes involved in the synthesis of toxins [12–15] are included. These virulence factors are typically involved in evolutionary arms races between plants and pathogens [16, 17].

Knowledge of the pathogenic cycle and that of virulence factors [18, 19] is crucial for designing effective crop protection strategies, including the development of resistant plant genotypes through classical plant breeding [20] or genetic engineering [21], fungicides [22], or the use of biological control strategies [23].

Studies of fungal pathogens and their interactions with plants have been performed using several approaches, from classical genetic, cell biology, and biochemistry [24–33], to the modern, holistic, and high-throughput omic techniques [34, 35] accompanied by proper bioinformatic tools [36]. In recent years, the study of fungal plant pathogens has been greatly promoted by the availability of their genomic sequences and resources for functional genomic analysis, including transcriptomics, proteomics, and metabolomics [37], which, in combination with targeted mutagenesis or transgenic studies, are unravelling molecular host-pathogen crosstalk, the complex mechanisms involving pathogenesis and host avoidance [38]. This review work makes an overview and summarizes the contribution of the most recent molecular techniques to the knowledge of phytopathogenic fungi biology and is mainly focused on the MS-based Proteomics approach.

2. From Structural to Functional Genomics

The importance of plant pathogenic fungi studies is underlined by the increasing number of fungal genome sequencing projects. Currently, over 40 fungal genomes

Table 2: Publicly available plant pathogenic fungal Genome sequences.

Phytopathogen Species[a]	URL
Ascomycota	
Dothydeomycetes	
Mycosphaerella fijensis (Banana black leaf streak)	http://genome.jgi-psf.org/Mycfi1/Mycfi1.home.html
Mycosphaerella graminicola (Wheat leaf blotch)	http://genome.jgi-psf.org/Mycgr1/Mycgr1.home.html
Pyrenophora tritici-repentis (Wheat disease)	http://www.broad.mit.edu/annotation/genome/pyrenophora_tritici_repentis/Home-html
Stagonospora nodorum (Wheat glume blotch)	http://www.broad.mit.edu/annotation/fungi/stagonospora_nodorum
	http://www.acnfp.murdoch.edu.au/Mission.htm
Eurotiomycetes	
Aspergillus flavus	http://www.aspergillusflavus.org/
Leotiomycetes	
Botrytis cinerea (Grape/other host grey rot) BO5.10	http://www.broad.mit.edu/annotation/fungi/botrytis_cinerea
T4	http://urgi.versailles.inra.fr/proyects/Botrytis/index.php
Sclerotinia sclerotium (Multi-host rot diseases)	http://www.broad.mit.edu/annotation/fungi/sclerotinia_sclerotium
Saccharomycetes	
Ashbya gossypii (Cotton/citrus fruits disease)	http://agd.vital-it.ch/index.html
Sordariomycetes	
Fusarium graminearum (Wheat/barley head blight)	http://www.broad.mit.edu/annotation/genome/fusarium_group
Fusarium oxysporum (Multi-host wilt disease)	http://www.broad.mit.edu/annotation/genome/fusarium_group
Fusarium verticillioides (Maize seed rot)	http://www.broad.mit.edu/annotation/genome/fusarium_group
Magnaporthe grisea (Rice blast disease)	http://www.broad.mit.edu/annotation/genome/magnaporthe_grisea/MultiHome.htlm
Nectria haematococca (Pea wilt)	http://genome.jgi-psf.org/Necha2/ Necha2.home.html
Verticillium dahliae VdLs.17 (Multi-host wilt)	http://www.broad.mit.edu/annotation/genome/verticillium_dahliae/MultiHome.html
Basidiomycota	
Pucciniomycetes	
Puccinia graminis (Cereal rusts)	http://www.broad.mit.edu/annotation/genome/puccinia_graminis
Ustilaginomycetes	
Ustilago maydis (Corn smut disease)	http://www.broad.mit.edu/annotation/fungi/ustilago_maydis

[a]Species are grouped by phylum and class. In parenthesis below the species's name and associated with each species, the most common or most widely recognized diseases are listed.

have been sequenced, 16 of which are phytopathogenic (Table 2), with more than 300 sequencing projects being in progress (Genomes Online database, http://www.genomesonline.org/). Sequence information, while valuable and a necessary starting point, is insufficient alone to answer questions concerning gene function, regulatory networks, and the biochemical pathways activated during pathogenesis. Based on the accumulation of a wealth of fungal genomic sequences, the traditional pursuit of a gene starting with a phenotype (forward genetics) has given way to the opposite situation where the gene sequences are known but not their functions. Thus, the challenge is now to decipher the function of the thousands of genes identified by genome projects, and reverse genetics methodologies are key tools in this endeavour [39].

TABLE 3: Original proteomics papers and reviews published on plant pathogenic fungi.

Fungus	Proteomic approach (reference)
Aspergillus ssp.	1-DE, MALDI-TOF-MS [40]
Aspergillus flavus	1-DE/2-DE, nanoLC-MS/MS [41]
	1-DE/2-DE, MALDI-TOF-MS [42]
Blumeria graminis f.sp. *hordei*	2-DE, MALDI-TOF/TOF-MS/MS [43]
	nanoLC-MS/MS [44]
	2-DE, MALDI-TOF/TOF-MS/MS, nanoLC-MS/MS, ESI-IT-MS/MS [45]
	2-DE, MALDI-TOF/TOF-MS/MS, ESI-IT-MS/MS [46]
Botrytis cinerea	2-DE, MALDI-TOF/TOF-MS/MS [47]
	1-DE, nanoLC MS/MS [48]
	1-DE, nanoLC MS/MS [49]
Curvularia lunata	2-DE, MALDI-TOF/TOF-MS/MS [50]
	1-DE/2-DE, nanoLC-Q-TOF-MS/MS [51]
	1-DE, CID-LTQ-MS [52]
Fusarium graminearum	2-DE, IT-MS/MS, iTRAQ-MS/MS [53]
	2-DE, ESI-MS/MS [54]
	Interactome [55]
Leptosphaeria maculans	1-DE, liquid-phase IEF, 2-DE [56]
Magnaporthe grisea	Interactome [57]
Phytophthora infestans	2-DE, MALDI-TOF-MS [58]
	2-DE, nanoLC-MS/MS [59]
Phytophthora palmivora	2-DE, MALDI-TOF-MS [58]
Phytophthora ramorum	HPLC-ESI-Q-TOF-nanoLC-MS/MS [60]
Pyrenophora tritici-repentis	2-DE, ESI-Q-TOF-MS/MS [61]
Pleorotus ostreatus	1-DE, MALDI-TOF-MS, ESI-Q-TOF-MS/MS/*de novo* sequencing [62]
Rhizoctonia solani	2-DE, MALDI-TOF-MS [63]
Sclerotinia sclerotiorum	2-DE, ESI-Q-TOF-nanoLC-MS/MS [64]
Stagnospora nodorum	2 -DE, LC-MS/MS [65]
	2-D LC-MALDI-MS/MS [66]
Uromyces appendiculatus	MudPit-MS/MS [67]
Ustilago maydis	2-DE, MALDI-TOF-MS, nanoLC-Q-TOF-MS/MS [68]

The study of gene function in filamentous fungi has made great advances in recent years [69]. Some of the techniques used in high-throughput reverse genetics approaches are targeted gene disruption/replacement (knock-out) [70], gene silencing (knock-down) [71], insertional mutagenesis [72], or targeting induced local lesions in genomes (TILLING) [73] (for review and examples see [39]). Thus, a number of pathogenicity factors have been targeted [74–76], and among them, several signaling pathways such as the cAMP and a mitogen activated protein kinase (MAPK) pathways have been shown to be crucial to virulence in several phytopathogenic fungi [77–82]. Random insertional mutagenesis is an excellent approach for dissecting complex biological traits, such as pathogenicity, because it does not require any prior information or assumptions on gene function. Recently, transposable elements (TEs) have been used for insertional mutagenesis and large-scale transposon mutagenesis has been developed as a tool for the genome-wide identification of virulence determinants in *F. oxysporum* [83].

Otherwise, transcriptomics, the global analysis of gene expression at the mRNA level, is also an attractive method for analyzing the molecular basis of fungal-plant interactions and pathogenesis [84–87]. For understanding the transcriptional activation or repression of genes during the infection process tools such as Differential Display (DD) [88], cDNA-Amplified Fragment-Length Polymorphism (cDNA-AFLP) [89], Suppression Subtractive Hybridization (SSH) [90], Serial Analyse of Gene Expression (SAGE) [91], expressed sequence tags (ESTs) [92], or DNA microarrays [91] have been developed in addition to older techniques such as Northern blotting, and they are reviewed in [84, 93].

3. Proteomics

Until the early 1990s most biological research was focused on the in vitro studies of individual components in which genes and proteins were investigated one at a time. This

strategy shifted in the early and mid 1990s to in vivo and molecular large-scale research, starting with structural genomics and transcriptomics research projects, then moving to proteomics, and recently to metabolomics. All these together constitute the methodological bases of the Modern Systems Biology [37, 94]. Since no single approach can fully clear up the complexity of living organisms, each approach does contribute and must be validated, this being considered as part of a multidisciplinary integrative analysis at different levels, extending from the gene to the phenotype through proteins and metabolites.

Within the "-omics" techniques, Proteomics constitutes, nowadays, priority research for any organism and configures a fundamental discipline in the postgenomic era. It is true that, at 2010, the realities are below the expectations originally generated and that the results gained over the last 15 years have shown that the dynamism, variability, and behaviour of proteins are more complex than had ever been imagined, especially as refers to a number of protein species per gene as a result of alternative splicing, reading frame, and posttranslational modifications, trafficking, and interactions, and considering that protein complexes, rather than individual proteins, are the functional units of the biological machines. However, and differently from other biological systems, mainly yeast [95] and humans [96], the full potential of proteomics is far from being fully exploited in fungal pathogen research, as refers to the low number of fungal pathogen species under investigation at the proteomic level, the low proteome coverage in those species investigated, and the almost unique use of classical, first generation techniques, those based on 2-DE coupled to MS.

The term proteomics was coined by Marc Wilkins, back during the 1994 Siena Meeting, to simply refer to the "PROTein complement of a genOME" [97]. Fifteen years later proteomics has become more than just an appendix of genomics or an experimental approach but a complex scientific discipline dealing with the study of the cell proteome. In the broadest sense, the proteome can be defined as being the total set of protein species present in a biological unit (organelle, cell, tissue, organ, individual, species, ecosystem) at any developmental stage and under specific environmental conditions. By using proteomics we aim to know how, where, when, and what for are the several hundred thousands of individual protein species produced in a living organism, how they interact with one another and with other molecules to construct the cellular building, and how they work with each other to fit in with programmed growth and development, and to interact with their biotic and abiotic environment. In the last ten years, excellent reviews and monographs on the fundamentals, concepts, applications, power, and limitations of proteomics have appeared [95, 98–104], some of them dealing with fungal pathogens [37, 105, 106]. It is not the objective of this review to comment or discuss every aspect but instead to show which one has been its contribution to the knowledge of fungal pathogens.

In Proteomics, several areas can be defined: (i) Descriptive Proteomics, including Intracellular and Subcellular Proteomics; (ii) Differential Expression Proteomics, (iii) Post-translational Modifications; (iv) Interactomics; and (v) Proteinomics (targeted or hypothesis-driven Proteomics). In the case of fungi, a new area can also be defined as Secretomics (the secretome is defined as being the combination of native proteins and cell machinery involved in their secretion), since many fungi secrete a vast number of proteins to accommodate their saprotrophic life-style; this would be the case of proteins implicated in the adhesion to the plant surface [107], host-tissue penetration and invasion effectors, [11, 108, 109], and other virulence factors [110].

Proteomics is constantly being renewed to respond to the question of the role of the proteins expressed in a living organism, experiencing, in the last ten years, an explosion of new protocols, and platforms with continuous improvements made at all workflow stages, starting from the laboratory (tissue and cell fractionation, protein extraction, depletion, purification, separation, MS analysis) and ending at the computer (algorithms for protein identification and bioinformatics tools for data analysis, databases, and repositories). These techniques will be briefly introduced and discussed in the next section.

Despite the technological achievements in proteomics, only a tiny fraction of the cell proteome has been characterized so far, and only for a few biological systems (human, fruit fly, *Arabidopsis*, rice). Even for these organisms, the function of quite a number of proteins remains to be investigated [99]. Proteomics techniques have a number of limitations, such as sensitivity, resolution, and speed of data capture. They also face a number of challenges, such as deeper proteome coverage, proteomics of unsequenced "orphan" organisms, top-down Proteomics [100], protein quantification [98], PTMs [105], and Interactomics [55, 57]. Most of these limitations and challenges reflect the difficulty of working with the biological diversity of proteins and their range of physicochemical properties.

The relevance of proteomics in plant fungal pathogens research is very well illustrated by the pioneer work on the *Cladosporium fulvum*-tomato interaction carried out by the Pierre de Wit research group back in 1985 [111] that allowed the characterization of the first avirulence gene product (Avr9) after purification from tomato apoplastic fluids by preparative polyacrylamide gel electrophoresis followed by reverse-phase HPLC and EDMAN N-terminal sequencing [112]. Later on, a number of avirulence gene product effectors have been discovered, mainly by genomic approaches [113]. Curiously, this pioneer work followed the typical proteomics strategy even before MS-based powerful techniques were developed. Another good example is the tomato-*F. oxysporum* pathosystem, in which the first effector of root invading fungi was identified and sequenced, in this case by MS, the Six1, corresponding to a 12 kDa cysteine-rich protein [114]. Other further protein effectors have been characterized in different fungi [115].

Next, we describe the state of the art in the methodology of fungal plant pathogen proteomics and summarize the works published in this field up to December 2009, which so far are 30 (Table 3) out of a total of over 14000 original papers in Proteomics in the last ten years.
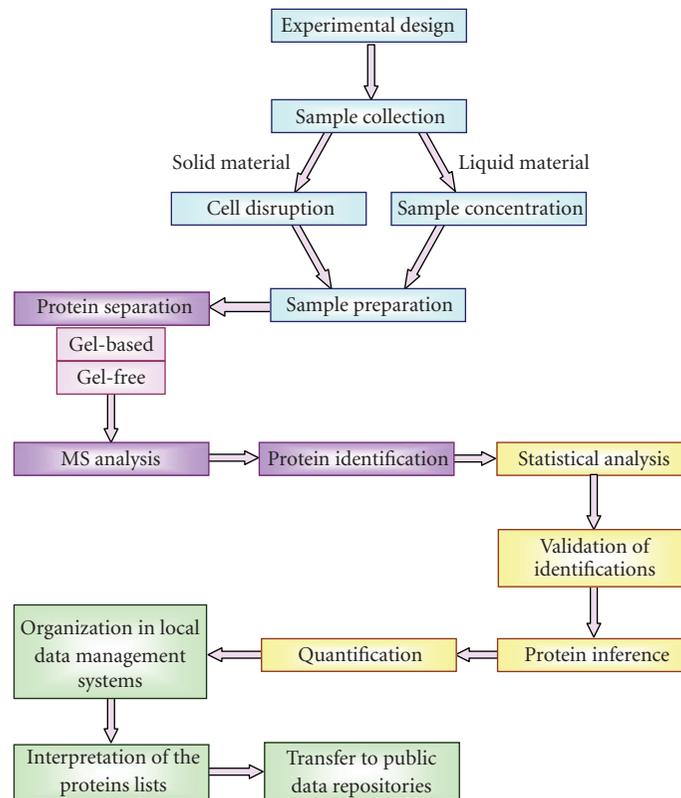
FIGURE 2: Schematic overview of the work flow in a fungal proteomatic approach (adapted from Deutsch et al., 2008 [116]).

3.1. Methodology. The workflow of a standard MS-based proteomics experiment includes all or most of the following steps: experimental design, sampling, tissue/cell or organelle preparation, protein extraction/fractionation/purification, labeling/modification, separation, MS analysis, protein identification, statistical analysis of data, validation of identification, protein inference, quantification, and data analysis, management, and storage (**Figure 2**) [102, 116]. The most appropriate protocol to be used depends on and must be optimized for the biological system (i.e., fungal species, plant species, organ, tissue, cells), as well as the objectives of the research (descriptive, comparative, PTMs, interactions, targeted Proteomics) [102].

3.1.1. Experimental Design. A good experimental design is crucial to the success of any proteomics experiment. Eriksson and Fenyö [117] have developed a simulation tool for evaluating the success of current designs and for predicting the performance of future, better-designed proteomics experiments. The simulation gives a holistic view of a general analytical experiment and attempts to identify the factors that affect the success rate. It has been used to predict the success of proteome analyses of Human tissue and body fluids that use various experimental design principles. Several parameters, are required to simulate the steps of a proteome analysis: (i) the distribution of protein amounts in the sample analyzed, (ii) the loss of analyte material and the maximal limit of the amount loaded at each step of

sample manipulation (e.g., separation, digestion, and chemical modification), (iii) the dynamic range, the detection limit, and the losses associated with MS analysis. Depending on what experiment is being modelled, the detection limit used in a simulation can represent either protein identification only (lower identification limit) or protein identification with quantification (lower quantification limit) [102].

The establishment of an adequate number of replicates is essential to any differential expression proteomics experiment. This number should be set up while taking into account the dynamic nature of the proteome, and a good number will allow a correct interpretation of the results and the confident assignment of any protein to the group of variable ones [102, 118]. Furthermore, a study of analytical and biological variability should be carried out. Thus, analytical variability examines both the experiment procedures (protein extraction, IEF, SDS-PAGE, gel staining-destaining) and the accuracy of the hardware and software in acquiring and analysing images, and biological variability tests look at several different samples. For example, in the studies to characterize the protein profile of the fungal phytopathogen *B. cinerea* [45], it previously determined the analytical and biological variability. Sixty-four major spots in three 2-DE replicate gels were analyzed and the average CVs were 16.1% for the analytical variance and 37.5% for the biological variance for this fungus. The analytical variance was similar to that reported for bacteria, plants, mice, or human extract and the average biological CV was

higher than that reported for other organisms grown under controlled conditions [45]. While characterizing the *Pinus radiata* needle proteome [119], we found differences in the standard error of mean spot quantity, depending on the number of replicates; the error ranged from 111 and 115 ng for two analytical and biological replicates to 58 and 59 ng for 10 analytical and 12 biological gel replicates.

Using more than six biological replicates did not significantly reduce the standard error; so this figure could be optimal for comparative proteomic experiments. Since normally this is not feasible, most papers in our literature review used only three biological replicates. Given the susceptibility of the data to variation, in 2-DE comparative proteomics it is necessary to be restrictive when deciding whether a spot showed variation. First, all the spots considered had to be consistent, that is, present or absent in all the biological replicates of the particular stage in question; second, when not qualitative (presence vs. absence), differences had to be statistically significant ($P \leq .05$, ANOVA); finally, the variance with respect to a control had to be higher than the average biological coefficient of variance determined for a representative set of at least 150 spots.

3.1.2. Cell Disruption and Sample Preparation. The importance of the extraction protocol in a proteomic experiment can be summarized in the following statement: only if you can extract and solubilize a protein, you have a chance of detecting and identifying it. This sentence sums up the importance of the preparation of protein samples in a proteomic experiment [102]. This is more important in the case of plant tissue or fungal material. In the case of filamentous fungi, the protein extraction is also influenced by the presence of a cell wall which makes up the majority of the cell mass. This cell wall is exceptionally robust [120], and the effective extraction of proteins is a critical step and essential for reproducible results in fungal proteomic studies. Therefore, cell breakdown is an important element in sample preparation for fungal proteomics.

Several early studies were performed to overcome this challenge by providing an effective means of cell lysis for an adequate release of intracellular proteins. For example, mechanical lysis was used to release the cytoplasmic protein via glass beads [59, 121–125], using a cell mill [68], or by sonication [40, 126, 127], these being more efficient than either chemical or enzyme extraction methods [128]. Shimizu and Wariishi [129] utilized an alternative approach to avoid the difficulty of lysing the fungal cell wall by generating protoplasts of *Tyromyces palustris*, since 2-DE patterns from protoplast were better visualized than protein obtained from disrupting the fungal cell wall using SDS extraction. However, the most widely used method for cell disruption is pulverizing the fungus material in liquid nitrogen using a mortar and pestle [45–47, 64, 126, 130, 131].

The extraction protocol most amply utilized implicates the use of protein precipitation media containing organic solvents, such as trichloroacetic acid (TCA)/acetone, followed by solubilization of the precipitate in an appropriate buffer. It allows an increase in the protein concentration

and removal of contaminants (salts, lipids, polysaccharides, phenols, and nucleic acids) which can be a problem during IEF [132] and prevents protease activities. This method has often been applied to the preparation of extracts from plants [133–135] and fungi [123, 136]. TCA-treatment makes subsequent protein solubilization for IEF difficult, especially with hydrophobic proteins. These problems have been partly overcome by the use of new zwitterionic detergents [137–142] and thiourea [143], or by a brief treatment with sodium hydroxide [123], which led to an increase in the resolution and capacity of 2-DE gels. Other protein extraction methods have reported an improvement when using acidic extraction solution to reduce streaking of fungal samples caused by their cell wall [144], as well as the use of a phosphate buffer solubilization before the precipitation [45, 46]. In a study to develop an optimized protein extraction protocol for *Rhizoctonia solani*, Lakshman et al. [63] compared TCA/acetone precipitation and phosphate solubilization before TCA/acetone precipitation. Both protocols worked well for *R. solani* protein extraction, although selective enrichment of some proteins was noted with either method. Finally, the combined use of TCA precipitation and phenol extraction gave a better spot definition, because it reduced streaking and led to a higher number of detected spots [47, 145].

In *B. cinerea*, our group has optimized a protein precipitation protocol from mycelium based on a combination of TCA/acetone and phenol/methanol protein extraction methods described by Maldonado et al. [135] with some modifications (Gonzalez-Fernandez et al., unpublished data). Mycelium is lyophilized, pulverized with liquid nitrogen using a mortar and pestle, sonicated in 10% (w/v) TCA/acetone solution with a sonic probe, washed sucessively with methanol and acetone, and finally a protein extraction is released with phenol/methanol precipitation method. We have used a similar protocol from conidia of this fungus by sonicating directly the spores collected in 10% (w/v) TCA/acetone solution (Gonzalez-Fernandez et al., unpublished results). In this sense, specific protocols for protein extraction from spores were optimized in *Aspergillus* ssp. using acidic conditions, step organic gradient, and variable sonication treatment (sonic probe and water bath) [40]. In this study, the use of a sonic probe was the best method to break the robust cell wall of conidia and obtain more proteins.

Special protocols are required in the case of secreted proteins, facing problems such as the very low protein concentration, sometimes below the detection limit of colorimetric methods for determining protein concentration such as Bradford, Lowry, or BCA, and the presence of polysaccharides, mucilaginous material, salts, and secreted metabolites (low-molecular organic acids, fatty acids, phenols, quinines, and other aromatic compounds). Moreover, the presence of these extracellular compounds may impair standard methods for protein quantification and can result in a strong overestimation of total protein [146]. This determination can also be affected by the high concentration of reagents from the solubilisation buffer (such as urea, thiourea or DTT) that may interfere in the spectrophotometric

measurement, producing an overestimation of the total amount of protein. Fragner et al., showed that, depending on the method, the differences varied in the order of two magnitudes, indicating that only the Bradford assay does not lead to an overestimation of the proteins [147].

Francisco and colleagues provided pioneering contributions to this field, establishing a sample preparation protocol for fungal secretome [148], including steps of lyophilization or ultrafiltration plus dialysis, precipitation (TCA/ethanol or chloroform/methanol), deglycosylation, and solubilization for SDS-PAGE or 2-DE.

Comparison of different standard methods for protein precipitation has demonstrated their limited applicability to analyzing the whole fungal secretome [149–152]. Usually, the fungal liquid culture is clarified by filtering and centrifuging, then dialyzed and concentrated up by lyophilizing. Recently, a new optimized protocol has been developed to obtain extracellular proteins from several higher basidiomycetes (*Coprinopsis cinerea*, *Pleorotus ostreatus*, *Phanerochaete chrysosporium*, *Polyporus brumalis,* and *Schizophyllum commune*) [147]. In this work, several precipitation methods, (TCA/acetone, phenol/methanol, other precipitation methods and an optimized protocol by high-speed centrifugation/TCA precipitation/Tris-acetone washes) were compared from liquid cultures of these fungi. The best method was the use of high-speed centrifugation, since it removed a considerable gelatinous material from the liquid culture supernatants. Then fungal proteins were effectively enriched by TCA precipitation and coprecipitated metabolites hampering 2-DE were removed through the application of Tris/acetone washes [147].

Vincent et al. [56], using the plant pathogenic fungus *Leptosphaeria maculans* and symbiont *Laccaria bicolor* grown in culture, have established a proteomic protocol for the analysis of the secretome. They evaluate different protocols, including ultrafiltration followed by TCA/acetone precipitation or phenol/ammonium acetate-phase partition, successive TCA/acetone precipitations without ultrafiltration, phenol/ether extraction without ultrafiltration, lyophilization, and prefractionation of secretome samples using liquid-phase IEF. Liquid-phase IEF followed by dialysis and lyophilization as a prefractionation prior to IPG-IEF produced the best results, with up to 2000 spots well resolved on 2-DE. This protocol can be applicable to a reduced number of samples and be very useful for descriptive proteomics but not for comparative proteomics experiments because of the excessive manipulation of the sample. Thus, in our work with *B. cinerea,* we aim to compare the proteome of wild-type and a high number of mutant strains (close to 100) affected in pathogenicity, and therefore we have optimized a protocol including the following steps: liquid media filtering, dialyzing, lyophilizing, and TCA/acetone-phenol/methanol protein precipitation. It is true that the number of resolved spots is much lower, but still enough for our purposes (Gonzalez-Fernandez et al., unpublished data).

A new field of study has been opened up with the analysis of infection structures such as appressorium and haustoria. In this case, specific protocols for the isolation of such structures are required [85] and the main problem is the large amount of material required for proteomic analysis if compared to transcriptomics. Godfrey et al. have developed a procedure for isolating haustoria from the barley powdery mildew fungus based on filtration and the use of differential and gradient centrifugation [44].

Up to now we have made reference to experiments with in vitro grown fungi; studies *in planta* are much more complicated due to the presence of both proteomes than of the plant and the pathogen. Bindschedler et al. [153] undertook a systematic shotgun proteomics analysis of the obligate biotroph *B. graminis* f. sp. *hordei* at different stages of development in the host: ungerminated spores, epiphytic sporulating hyphae, and haustoria, this being, as far as we know, the only large-scale comparative study of proteomes of phytopathogenic fungi during *in planta* colonization in addition to those analyses of whole extracts from host infected tissue [154, 155] or intercellular washing fluids [156].

In short, since no single protein extraction protocol can capture the full proteome, the chosen protocol should be optimized for the research objective. The ideal method should be highly reproducible and should extract the greatest number of protein species, while at the same time reduce the level of contaminants and minimize artifactual protein degradation and modifications [148, 157, 158]. Fortunately, each protocol takes us to a specific fraction of the proteome, thus being complementary [135]. Another issue to consider is the extreme complexity of the proteome and the wide dynamic range in protein abundance, which exceeds the capability of all currently available analytical platforms. Sample prefractionation is a good approach to reducing the complexity of the proteome sample and decreasing the dynamic range [159], with EQUALIZER being the last developed technology [160].

3.1.3. Protein Separation, Mass Spectrometry and Protein Identification. In fungal plant pathogen research electrophoresis, including denaturing 1-DE SDS-PAGE and 2-DE, with IEF as first dimension, and SDS-PAGE as the second, is almost the only protein separation technique employed, with both crude total protein extracts and protein fractions obtained from various prefractionation procedures [161]. Despite its simplicity, 1-DE can still be quite a valid technique providing relevant information, especially in the case of comparative proteomics with large numbers of samples to be compared. Thus, using this technique, it is possible to distinguish between genotypes of different wild-type strains of *B. cinerea* and identify proteins involved in the pathogenicity mechanisms (Figure 3) (Gonzalez-Fernandez et al., unpublished results). With appropriate software, 1-DE is a simple and reliable technique for finger-printing crude extracts and it is especially useful in the case of hydrophobic and low-molecular-weight proteins [162]. The combination of 1-DE, band cutting, trypsin digestion, and LC separation of the resulting peptides remains the proteomic technique capable of providing the greatest protein coverage [163, 164]. Therefore, the 1-DE is a good approach to obtaining preliminary results before the study by 2-DE
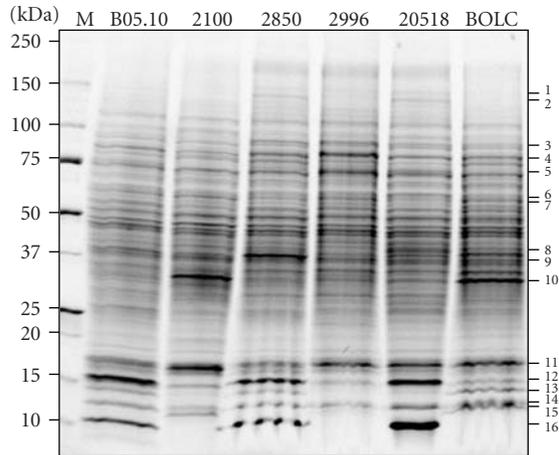
FIGURE 3: One-DE of 15 μg of mycelium protein extract of six different strains of *B. cinerea* (B05.10, CECT 2100, CECT 2850, CECT 2996, CECT 20518, BOLC (isolated from infected lentil plants)). This approach allowed us to observe differences in the protein band patterns among strains. The bands were cut out and the protein identification was made using MALDI-TOF/TOF MS/MS, and PMF search and a combined search (+MS/MS) were performed in nrNCBI database of proteins using MASCOT. Some of these proteins identified have been reported to be involved in pathogenicity in *B. cinerea* or in other phytopathogenic fungi, such as malate dehydrogenase (10), woronin body major protein (11), peptidyl-prolyl cis-trans isomerase (14) and PIC5 protein (15), or implicated in fungal growth and differentiation, such as nucleoside diphosphate kinase (12). The abundance of these proteins was different among isolated (Gonzalez-Fernandez et al., unpublished results).

[41, 42, 50, 165]. For example, the use of 1-DE in combination with MS/MS analysis has allowed the detection of both known [166] or new proteins [62] of interest in fungal pathogenicity.

In fungal proteomics, 2-DE separation techniques [132, 167, 168] are widely used (Table 3), these having the advantage of separating the proteins at the protein species level with a high resolution of up to 10,000 spots [169]. Briefly, the 2-DE consists of a tandem pair of electrophoretic separations: in the first dimension, proteins are resolved according to their isoelectric points (pIs) normally using IEF, and in the second dimension, proteins are separated according to their approximate molecular weight using SDS-PAGE. These proteins can then be detected by a variety of staining techniques: (i) organic dyes, such as colloidal Coomassie blue staining, (ii) zinc-imidazole staining, (iii) silver staining, and (iv) fluorescence-based detection. Excellent reviews describe and discuss the features and protocols of electrophoretic separations in proteomics strategies [132, 170]. The main advantages of 2-DE are high protein separation capacity and the possibility of making large-scale protein-profiling experiments. However, the reproducibility and resolution of this technique are still remaining challenges. Moreover, this method was reported to under-represent proteins with extreme physicochemical properties (size, isoelectric point, transmembrane domains) and those of a low abundance

[171]. These limitations to analytical protein profiling have led to the more recent development of techniques based on LC separation of proteins or peptides, including two-dimensional liquid-phase chromatography (based on a high-perfomance chromatofocusing in the first dimension followed by high-resolution reversed-phase chromatography in the second one) [172], and one-dimensional 1-DE-nanoscale capillary LC-MS/MS, namely, GeLC-MS/MS (this technique combines a size-based protein separation with an in-gel digestion of the resulting fractions). Recorbert et al. [173] explored the efficiency of GeLC-MS/MS to identify proteins from the mycelium of *Glomus intraradices* developed on root organ cultures, reporting on the identification of 92 different proteins. Overall, this GeLC-MS/MS strategy paves the way towards analysing on a large-scale fungal response environmental cues on the basis of quantitative shotgun protein profiling experiments.

Despite the existence of quite a number of different methods developed for protein extraction and separation, it is clear that, all in all, it is not enough to allow for the analysis of entire proteomes (organelle, cell, tissue, or organ). Some methodologies have proven to be more powerful and decisive than others, with regard to the number of proteins identified. This is the case of Multidimensional Protein Identification Technology (MudPIT), an LC-based strategy, which allows the detection of a much larger number of proteins compared to gel-based methods, its drawback being the lack of quantitative data [174]. MudPIT was first applied to the fungal proteome of the *S. cerevisiae* and yielded the largest proteome analysis to date, in which a total of 1484 proteins were detected and identified [175]. The categorization of these hits demonstrated the ability of this technology to detect and identify proteins rarely seen in proteome analysis, including low abundance proteins like transcription factors and protein kinases [175]. It has been reported that a set of proteins can only be detected by a specific technology [176, 177], which is in agreement with the idea that a combination of different methodologies is still needed to characterize entire proteomes [131]. MudPIT has been used to compare the proteome from germinating and ungerminated asexual uredospores of the biotrophic fungal pathogen *Uromyces appendiculatus*, which is the causal agent of rust disease in beans [67].

Mass spectrometry is the basic technique for global proteomic analysis due to its accuracy, resolution, and sensitivity, small amounts of sample (femtomole to attomole concentrations), and having the capacity for a high throughput. It allows not only to profile a proteome from a qualitative and quantitative point of view but also, and more important, to identify protein species and characterize postranslational modifications. Proteins are identified from mass spectra of intact proteins (top-down proteomics) or peptide fragments obtained after enzymatic (mostly tryptic peptides) or chemical treatment (bottom-up proteomics). Protein species are identified by comparison of the experimental spectra and the theoretical one obtained in silico from protein, genomic, ESTs sequence, or MS spectra databases. For that purpose, different instrumentation, algorithms, databases, and repositories are available [178, 179].

Different strategies and algorithms can be used for protein identification (i) including peptide mass fingerprinting (PMF, only valid if the protein sequence is present in the database of interest, generally used if the genome of the organism of study is fully sequenced); (ii) tandem mass spectrometry (where peptide sequences are identified by correlating acquired fragment ion spectra with theoretical spectra predicted for each peptide contained in a protein sequence database, or by correlating acquired fragment ion spectra with libraries of experimental MS/MS spectra identified in previous experiments); (iii) *de novo* sequencing, where peptide sequences are explicitly read out directly from fragment ion spectra; and (iv) hybrid approaches, such as those based on the extraction of short sequence tags of 3–5 residues in length, followed by "error-tolerant" database searching [179]. In a genomic-based proteomics strategy the percentage of proteins identified from MS data is dependent on the availability of genomic DNA or EST sequences. The construction of protein repositories with signature peptides and derived MS spectra will open up new possibilities for protein identification. The availability of ESTs from unsequenced "orphan" organisms as is the case of most plant fugal pathogens of interest will increase the percentage of identified proteins [93]. There are relatively few and a slow accumulation of EST data derived from a number of plant fungal pathogens and related species in public databases (e.g., dbEST at the National Centre for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/dbEST/; and COGEME, Phytopathogenic Fungi and Oomycete EST Database [180, 181] at Exeter University, UK, http://cogeme.ex.ac.uk/). In Table 4, the number of EST entries for a number of fungi at the NCBI and Dana Faber databases is listed.

3.1.4. Second-Generation Techniques for Quantitative Proteomics. By using proteomics it is not only aimed to identify protein species (main objective of descriptive proteomics) but also quantify them, at least in relative terms, by comparing two biological units (genotypes, cells, organules) under different spatiotemporal parameters and environmental conditions. Absolute, rather than relative, protein quantitation remains one of the main challenges in proteomics [102]. There are different methods to dissect the proteome in a quantitative manner: (i) methods based on 2-DE with poststaining, such as colloidal Comassie blue staining [182] and fluorescence staining [183], or prelabeling such as two-dimensional fluorescence difference gel electrophoresis (2-DIGE) [184, 185]; and (ii) gel-free methods based on in vitro or in vivo protein targetting with a stable isotope, such as isotope-coded affinity tags (ICATs) [186, 187], ^{18}O labeling [188], or stable isotope labelling in cell culture (SILAC) [189], or isobaric tags, such as isobaric tag for relative and absolute quantitation (iTRAQ) [190]. Novel, label-free approaches, such as spectral counting, are being developed [191, 192]. The use of second generation proteomic techniques based on protein labelling and those label-free ones are far from being fully exploited in fungal pathogen research.

TABLE 4: Number of ESTs entries for some fungi of interest up to December 2009.

Fungus	Dana Faber/NBCI[a]
Aspergillus flavus	20372/22452
Aspergillus niger	No entries/47082
Blumeria graminis hordei	No entries/17142
Botrytis cinerea	No entries/28531
Claviceps purpurea	No entries/8789
Curvularia lunata	No entries/1488
Fusarium graminearum	No entries/58011
Fusarium oxysporum	No entries/17478
Fusarium verticillioides	86908/87134
Leptosphaeria maculans	No entries/20034
Magnaporthe grisea	87403/110613
Puccinia graminis	No entries/209
Phytophthora infestans	90287/164143
Phytophthora palmivora	No entries/14824
Pyrenophora tritici-repentis	No entries/1
Nectria haematococca mpVI	No entries/33122
Sclerotinia sclerotiorum	No entries/2578
Stagonospora nodorum	No entries/16447
Ustilago maydis	No entries/39717

[a]Data taken from The Gene Index Proyect at the Dana Farber Cancer Institute (http://compbio.dfci.harvard.edu/tgi/plant.html) and NBCI (http://www.ncbi.nlm.nih.gov/).

In 2-DIGE, proteins in two samples are labeled in vitro through cysteine or lysine residues using two different fluorescent cyanine dyes differing in their excitation and emission wavelengths, but with an identical relative mass. Labeled samples are then mixed and subjected to 2-DE on the same gel. After consecutive excitation with both wavelengths, the images are overlaid and normalized, whereby only differences between the two samples are visualized [193]. 2-DIGE enables to perform high-throughput, differential protein expression analysis to compare directly, on a single gel, the differences in protein expression levels between different complex protein samples. The main advantage of 2-DIGE on 2-DE is its unrivalled performance, attributable to a unique experiment, in which each protein spot on the gel is represented by its own internal standard [105].

The classical proteomic quantification electrophoretic methods utilizing dyes, fluorophores, or radioactivity have provided very good sensitivity, linearity, and dynamic ranges, but they suffer from two important shortcomings: first, they require high-resolution protein separation typically provided by 2-DE gels, which limits their applicability to abundant and soluble proteins; and second, they do not reveal the identity of the underlying protein [194]. Both of these problems are overcome by modern LC-MS/MS techniques [95, 195–197]. However, the MS-based techniques are not inherently quantitative because proteolytic peptides exhibit a wide range of physicochemical properties (size, charge, hydrophobicity) that lead to large differences in mass spectrometric response. Therefore, in MS-based gel-free quantification it is necessary

to use isotopic labeling. Observed peak ratios for isotopic analogs are highly accurate, because there are no chemical differences between the species, and they are analyzed in the same experiment. Mass spectrometry can recognize the mass difference between the labeled and unlabeled forms of a peptide and the quantification is achieved by comparing their respective signal intensities [194].

A number of isotopic labeling techniques have recently been proposed that share the requirement of the chemical modification of the peptides or proteins. One of these strategies is the ICAT method for relative quantitation of protein abundance [186]. In this approach, an isotopically labeled affinity reagent is attached to particular amino acids in all proteins in the population. After digestion of the protein to peptides, as a necessary step in all mainstream proteomic protocols, the labeled peptides are affinity-purified using the newly incorporated affinity tag, thereby achieving a simplification of the peptide mixture at the same time as incorporating the isotopic label. This method addresses many of the above limitations and leads to a larger number of identifications of cysteine-containing peptides. However, the method is performed by cross-linking peptides to beads via their cysteine groups and photo-releasing them afterwards, which may compromise low-level analysis. The iTRAQ is used to identify and quantify proteins from different sources in one experiment. The method is based on the covalent labeling of the N-terminus and side-chain amines of peptides from protein digestions with tags of varying mass. The fragmentation of the attached tag generates a low molecular mass reporter ion that can be used to relatively quantify the peptides and the proteins from which they originated [198]. In SILAC (stable isotope labeling by amino acids in cell culture), labeled, essential amino acids are added to amino acid-deficient cell culture media and are therefore incorporated into all proteins as they are synthesized, "encoded into the proteome" [189]. No chemical labeling or affinity purification steps are performed, and the method is compatible with virtually all cell culture conditions. Finally, label-free protein quantification methods are promising alternatives. It is based on precursor signal intensity, which is, in most cases, applied to data acquired on high mass precision spectrometers. The mass spectral peak intensities of the peptide ion correlate well with protein abundances in complex samples [199–201]. Another label-free method is spectral counting, which simply counts the number of MS/MS spectra identified for a given peptide and then integrates the results for all measured peptides of proteins that are quantified [202]. An advantage of this technique is that relative abundances of different proteins can in principle be measured. These new quantification techniques have become powerful tools to overcome the inherent problems of the 2-DE including identification of proteins of a low abundance, high hydrophobicity, extreme pl or high MW.

These second-generation MS technologies for Quantitative Proteomics have not begun to be applied to Fungal Proteomics yet. The only one nongel-based quantitative proteomics example is the use of iTRAQ to study the profile protein expression differences on *F. graminearum*

which allowed the identification of numerous candidate pathogenicity proteins [53].

Although this review is focused on MS-based proteomics, we want to make a brief mention of protein microarrays because they are powerful tools for individual studies as well as systematic characterization of proteins and their biochemical activities and regulation [203]. The arrays can be used to screen nearly the entire proteome in an unbiased fashion and have an enormous utility for a variety of applications. These include protein-protein interactions, identification of novel lipid- and nucleic acid-binding proteins, and finding targets of small molecules, protein kinases, and other modification enzymes.

In short, all these technologies have a great potential in protein separation and remain a challenge for future research works in Fungal Proteomics.

3.1.5. Data Analysis and Statistical Validation. Proteomics tools generate an important amount of data, because a single proteomics experiment reveals the expression information for hundreds or thousands of proteins. Therefore, data analysis and bioinformatics are essential for this type of research and in many cases take more time than the actual experiment and require special skills and tools (for review see [116, 118]). All 2-DE software permits a fast and reliable gel comparison, and multiple gel analyses, including filtering of 2-DE images, automatic spot detection, normalization of the volume of each protein spot, and differential and statistical analyses [204–206]. A great resource for finding software tools for proteomics can be found in the website https://proteomecommons.org/.

Before the protein identification, the remaining challenge is to determine whether the putative identification is, in fact, correct. Statistical tools help us to validate information. Although postsearch statistical validation still does not enjoy universal application, its importance has been recognized by most researchers and codified in the editorial policies of some leading journal [207]. In this decade, an important number of commercial software involving even more powerful algorithms and statistical tools than those of the previous generations have been designed to help researchers deal with the sheer quantity of data produced [194, 195, 208].

Statistical data analyses can be classified as univariated or multivariated [209]. The univariated methods, such as the Student's *t*-Test, are used to detect significant changes in the expression of individual proteins. They are the simplest to interpret conceptually and the most common ones used. The multivariated methods, such as principle component analysis (PCA), look for patterns in expression changes and utilize all the data simultaneously. Early expression studies compared one sample with another, generally by the calculation of a ratio, and the analyses were restricted to looking for changes above a threshold determined by the system experimental noise. This method of analysis limits the sensitivity of the system, as biologically relevant changes smaller than the threshold cannot be detected. Using a threshold is a rather simplistic approach and does not take into account the variability of each protein, running the risk of selecting

TABLE 5: Useful online resources and Fungal Genome and Proteome Databases.

Name/description	URL
Genome Databases	
National Center for Biotechnology Information (NCBI).	http://www.ncbi.nlm.nih.gov/
NIH genetic sequences database.	http://www.ncbi.nlm.nih.gov/Genbank/
Fungal Genomes Central, information and resources pertaining to fungi and fungal sequencing projects.	http://www.ncbi.nlm.nih.gov/projects/genome/guide/fungi/
The Gene Index proyect (GI). The Computational Biology and Functional Genomics Laboratory, and the Dana-Faber Institute and Public School of Public Health.	http://compbio.dfci.harvard.edu/tgi/fungi.html
Fungal Genome Initiative of The Broad Institute (FGI).	http://www.broadinstitute.org/science/projects/fungal-genome-initiative/fungal-genome-initiative
Genoscope, Sequencing National Centre.	http://www.genoscope.cns.fr/spip/Fungi-sequenced-at-Genoscope.html
Joint Genome Institute (JGI).	http://www.jgi.doe.gov/
The Genome Center at Washington University (WU-GSC).	http://genome.wustl.edu/genomes/list/plant_fungi
The Sanger Institute fungal sequencing.	http://www.sanger.ac.uk/Projects/Fungi/
Genome projects.	http://genomesonline.org/
The MIPS *F. Graminearum* Genome Database.	http://mips.gsf.de/projects/fungi/Fgraminearum/
The MIPS *U. Maydis* Database.	http://mips.gsf.de/genre/proj/ustilago
The MIPS *Neurospora crassa* Genome Database.	http://mips.helmholtz-muenchen.de/genre/proj/ncrassa/
COGEME, Phytopathogenic Fungi and Oomycete EST Database (v1.6), constructed and maintained by Darren Soanes (University of Exeter, UK).	http://cogeme.ex.ac.uk/
SGD, *Saccaromyces* Genome Database, scientific database of the molecular biology and genetics of the yeast *Saccharomyces cerevisiae*.	http://www.yeastgenome.org/
e-Fungi, warehouse which integrates sequence data (genomic data, EST data, Gene Ontology annotation, KEGG pathways and results of the following analyses performed on the genomic data) from multiple fungal sequences in a way that facilitates the systematic comparative study of those genomes (School of Computer Science and the Faculty of Life Sciences at the University of Manchester and the Departments of Computer Science and Biological Sciences at the University of Exeter).	http://beaconw.cs.manchester.ac.uk/efungi/execute/welcome
CADRE, Central Aspergillus Database Repository, resource for viewing assemblies and annotated genes arising from various Aspergillus sequencing and annotation projects.	http://www.cadre-genomes.org.uk/
FungalGenome, website with several links and references for the currently available fungal genomes sequences or proposed fungal genomes.	http//:fungalgenomes.org/wiki/Fungal_Genomes_Links
Proteome Databases	
The Expasy (Expert Protein Analysis System) proteomics server of Swiss Institute of Bioinformatics (SIB). Analysis of protein sequences, structures and 2-D-PAGE.	http://ca.expaxy.org/
MIPS, Munich Information Center for Protein Sequences.	http://mips.gsf.de/
The PRIDE, Proteinomics IDEntifications Database. EMBL-EBI (European Bioinformatic Institute).	http://www.ebi.ac.uk/pride/

TABLE 5: Continued.

Name/description	URL
Integr8, Integrated information about deciphered genomes and their corresponding proteomes. EMBL-EBI.	http://www.ebi.ac.uk/integr8/EBI-Integr8-HomePage.do
SNAPPVIEW (Structure, iNterfaces and Aligments for Protein-Protein Interactions).	http://www.compbio.dundee.ac.uk/SNAPPI/download.jsp
Phospho3. Database of three-dimensional structures of phosphorylation sites.	http://cbm.bio.uniroma2.it/phospho3d/
Proteome Analyst PA-GOSUB 2.5. Sequences, predicted GO molecular functions and subcellular localisations.	http://www.cs.ualberta.ca/~bioinfo/PA/GOSUB/
RCSB, The Research Collaboratory for Structural Bioinformatics. Protein Database (PDB).	http://www.rcbs.org/pdb/home/home.do
PDB-Site. Comprehensive structural and functional information on PTMs, catalytic active sites, ligand binding (protein-protein, protein-DNA, protein-RNA interacions) in the Protein Data Bank (PDB).	http://wwwmgs.bionet.nsc.ru/mgs/gnw/pbbsite/
WoLF PSORT, Protein Subcellular Localization Prediction.	http://wolfpsort.org/
NMPdb, Nuclear Matrix Associated Proteins.	http://cubic.bioc.columbia.edu/db/NMPdb/
TargetP, predicts the subcellular location of eukaryotic proteins, based on the predicted presence of the N-terminal presequences.	http://www.cbs.dtu.dk/services/TargetP/
MitoP2, Mitochondrial Database. This database provides a comprehensive list of mitochondrial proteins of yeast, mouse, *Arabidopsis thaliana*, neurospora and human.	http://www.mitop.de:8080/mitop2/
The SecretomeP, Prediction of protein secretion and information on various PTMs and localisational aspect of the protein.	http://www.cbs.dtu.dk/services/SecretomeP/
MASCOT, a powerful search engine that uses MS data to identify proteins from primary sequence databases.	http://www.matrixscience.com/
VEMS, Virtual Expert Mass Spectrometrist. Program for integrated proteome analysis.	http://www.yass.sdu.dk/
The NetPhos server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.	http://www.cbs.dtu.dk/services/NetPhos/
ProPrInt, Protein-Protein Interaction Predictor. Compilation of web-resources in the field of Protein-Protein Interaction (PPI).	http://www.imtech.res.in/raghava/proprint/resources.htm
ProteomeCommons, public proteomics database for annotations and other information linked to the Tranche data repository and to other resources. It provides public access to free, open-source proteomics tools and data.	https://proteomecommons.org
Fungal Proteome Specialized Databases	
MPID, Protein-protein interaction Database of *M. grisea*.	http://bioinformatics.cau.edu.cn/cgi-bin/zzd-cgi/ppi/mpid.pl
FPPI, Protein-protein interaction database of *F. graminearum*.	http://csb.shu.edu.cn/fppi

variable proteins due to sample selection. The use of a fold change has a potential role in preliminary experiments but the limitation of this method must always be considered when interpreting the data [209].

Hypothesis tests, for example, Student's t-Test, assess whether the differences between groups are an effect of chance arising from a sampling effect or reflect a real statistical significant difference between groups [209]. Hypothesis tests are usually stated in terms of both a condition that is in doubt (the null hypothesis) and a condition that is believed to exist (the alternative hypothesis). The tests calculate a p-score, which is the probability of obtaining these results assuming that the null hypothesis is correct. Hypothesis tests can be divided into two groups: parametric and nonparametric. Parametric tests assume that the distribution of the variables being assessed belong to known probability distributions. For example, the Student's t-Test assumes that the variable comes from a normal distribution. Nonparametric tests, also called distribution-free methods, do not rely on estimation of parameters such as the mean. An example of nonparametric tests is the Mann-Whitney U-Test, which ranks all values from low to high and then compares the mean ranking the two groups. Otherwise, Student's t-Test or Mann-Whitney U-Test compares two groups and ANOVA or Krustal-Wallis compares more than two groups. These tests analyze individual spots instead of the complete set, omitting information about correlated variables. Biron et al. recommend assessing the normality for each protein species and then selecting either a parametric or nonparametric test [208].

In expression studies, many thousands of statistical tests are conducted, one for each protein species. A substantial number of false positives may accumulate which is termed the multiple testing problem and is a general property of a confidence-based statistical test when applied many times [209]. One approach to addressing the multiple testing problem is to control the family wise error rate (FWER), which control the probability of one or more false rejections among all the tests conducted. The simplest and most conservative approach is the Bonferroni correction, which adjusts the threshold of significance by dividing the per comparison error rate (PCER) by the number of comparisons being completed [210]. This has led to the application of methodologies to control the false discovery rate (FDR), where the focus is on achieving an acceptable ratio of true- and false-positives. The FDR is a proportion of changes identified as significant that are false [211]. An extension to the FDR calculates a Q-value for each tested feature and is the expected proportion of false positives incurred when making a call that this feature has a significant change in the expression [212]. For each P-value, a Q-value will be reported on an overall estimation for the proportion of species changing in the study.

Multiple testing correction methods, such as the Bonferroni correction and testing for the false discovery rate (FDR) [213], fit the Student t-test or ANOVA values for each protein spot to keep the overall error rate as low as possible. Multivariate data analysis methods, such as PCA, are now used to pinpoint spots that differ between samples.

These multivariate methods focus not only on differences in individual spots but also on the covariance structure between proteins [214]. However, the results of these methods are sensitive to data scaling, and they may fail to produce valid multivariate models due to the large number of spots in the gels that do not contribute to the discrimination process [215]. One of the limitations of PCA analysis is that it does not allow to miss values, a problem that can be avoided by imputing them when possible (if enough replicates are available) [216].

3.1.6. Databases and Repositories. The huge amount of data generated are being deposited and organized in several databases available to the scientific community: the UniProt knowledgebase reported by Schneider et al. [217] and other Proteome Databases mentioned in Table 5. After 20 years of Proteomics research, it is possible to look back at previous research and publications, identifying errors from the experimental design, the analysis, and the interpretation of the data [179]. In addition, data validation is done in a purely descriptive or speculative manner, as well as it is common to find low-confidence protein identification in the literature, especially in the case of unsequenced organisms and inappropriate statistical analyses of results have often been performed. It is interesting to see how many manuscripts contain the term "proteome" when probably only a tiny fraction of the total proteome has been analysed. About this problem, HUPO's Proteomic Standard Initiative has developed guidance modules [218] that have been translated into Minimal Information about a Proteomic Experiment (MIAPE) documents. The MIAPE documents recommend proteomic techniques that should be considered and followed when conducting a proteomic experiment. Proteomics journals should be, and in fact are, extremely strict when recommending that investigators follow the MIAPE standards for publishing a proteomic experiment. On the other hand, many journals recommend or require the original data generated in a proteomic experiment to be submitted to public repositories [207, 219]. A shift in the protein identification paradigm is currently underway, moving from sequencing and database searching to spectrum searching in spectral libraries. This underscores the importance of repositories for Proteomics [220–222]. The main public peptide and protein identification repositories are GPMDB (Global Proteome Machine database) [223], PeptideAtlas [224], and Proteomics IDEntifications database (PRIDE) [225]. Other emerging and smaller systems include Genome Annotating Proteomic Pipeline database (GAPP database) [226], Tranche (Falkner, J. A., Andrews, P. C., HUPO Conference 2006, Long Beach, USA, Poster presentation), PepSeeker [227], Max-Planck Unified database (MAPU) [228], the Open Proteomic Database (OPD) [229], and the Yeast Resource Center Public Data Repository (YRC PDR) [230].

3.2. Proteomics of Plant Pathogenic Fungi. Several proteomic studies have been carried out in order to understand fungal pathogenicity or plant-fungus interactions

TABLE 6: Original proteomics papers published in plant-pathogenic fungi interactions.

Pathogen-Host	Description of study (References)
Alternaria bassicicola-Arabidopsis	Study of change in the *Arabidopsis* secretome in response to salicylic acid and identifying of several proteins involved in pathogen response such as GDSL LIPASE1 (GLIP1) [231].
Aphanomyces eutiches-Medicago truncatula	Identification of several proteins which play a major role during root adaptation to various stress conditions [232], and study of parasitic plant-pathogen interactions formed between legumes and this oomycete [233, 234].
Black point disease-Barley	Identification of a novel late embryogenesis abundant (LEA) protein and a barley grain peroxidase 1 (BP1) that were specifically more abundant in healthy grain and black pointed grain, respectively [235].
B. graminis hordei-Barley	Systematic shotgun proteomics analysis at different stages of development of powdery mildew in the host to gain further understanding of the biology during infection of this fungus [153].
Cladosporium fulvum-Tomato	Identified 3 novel fungal secretory proteins [236].
Cronartium ribicola-Pinus strobus	Study of molecular basis of white pine blister rust resistance [237].
Diploidia scrobiculata-Pinus nigra Sphaeropsis sapinea-Pinus nigra	Study about defense protein responses in phloem of Austrian pine inoculated with *D. scrobiculata* and *S. sapinea* [238].
Erysiphe pisi-Pea	Identification of proteins implicated in powdery mildew resistance [239].
F. graminearum-Barley	Identification of proteins associated with resistance to *Fusarium* head blight in barley [240].
F. graminearum-Wheat	Identification of proteins associated with resistance to *Fusarium* head blight in wheat [241] and which have a role in interaction between *F. graminearum* and *T. aestivum* [155].
F. graminearum-Wheat	Identification of proteins associated with resistance to scab in wheat spikes [238].
F. moniliforme-Arabidopsis	Study of changes in the extracellular matrix of *A. thaliana* cell suspension cultures with fungal pathogen elicitors of *F. moniliforme* [242].
F. oxysporum-Sugar beet	Study of resistance to *F. oxysporum* disease [243].
F. oxysporum-Tomato	Identification of 21 tomato and 7 fungal proteins in the xylem sap of tomato plants infected by *F. oxysporum* [156].
F. verticillioides-Maize	Identification of protein change patterns in germinating maize embryos in response to infection with *F. verticillioides* [244].
Fusarium-Arabidopsis	Identification of differentially expressed proteins in response to treatments with pathogen-derived elicitors to identify pivotal genes' role in pathogen defence systems [245].
Fusarium-Maize	Study of the role of the extracellular matrix in signal modulation during pathogen-induced defence responses [246].
Gossypium hirsutum-Cotton	Identified pathogen-induced cotton proteins implicated in post-invasion defence reponses (PR-proteins related to oxidative burst), nitrogen metabolism, amino acid synthesis and isoprenoid synthesis [247].
Hyaloperonospora parasitica-Arabidopsis B. cinerea-Arabidopsis	Study of pathogenic resistance of Arabidopsis wild-type and CaHIR1-overexpressing transgenic plants inoculated with these fungi, among other pathogens [248].

TABLE 6: Continued.

Pathogen-Host	Description of study (References)
Leptosphaeria maculans-Brassica	Identification of *Brassica* proteins involved in resistance to this fungus [249].
L. maculans-Brassica carinata L. maculans-Brassica napus	Study of changes in the leaf protein profiles of *Brassica napus* (highly susceptible) and *Brassica carinata* (highly resistant) in order to understand the biochemical basis for the observed resistance to *L. maculans* [250].
*M. grisea-*Rice	Change protein analysis during blast fungus infection of rice leaves with different levels of nitrogen nutrient [251]. Analysis of differentially expressed proteins induced by blast fungus in suspension-cultured cells [252] in leaves [253] and during appressorium formation [254]. Proteomic approach of differentially expressed proteins in rice plant leaves at 12 h and 24 h after treatment with the glycoprotein elicitor CSB I, purified from ZC(13), a race of the rice blast fungus *M. grisea* [255].
Marsonina brunnea f. sp. *Multigermtub-Populus euramericana*	Identification of proteins related to black spot disease resistance in poplar leaves [256].
*Moliniophtora perniciosa-*Cocoa	Optimization of protein extraction for cocoa leaves and meristemes infected by this fungus that causes witches' broom disease [257].
*Nectria haematococca-*Pea	Study of extracellular proteins in pea roots inoculated with *N. haematococca* [258].
*Penincillium exposum-Piccia membranefaciens-*Peach fruit	Peach fruit inoculated with *P. exposum* and treated with SA and *P. membranefaciens* [259].
Phellinus sulphurasces-Pseudotsuga menzeii	Comparative proteomic study to explore the molecular mechanisms that underlie the defense response of Douglas-fir to laminated root rot disease caused by *P. sulphurascens* [260].
*Peronospora viciae-*Pea	Catalogued host (pea) leaf proteins, which showed alternation in their abundance levels during a compatible interaction with *P. viciae* [261].
Plasmodiophora brassicae-Brassica napus	Study of changes in the root protein profile of canola with clubroot disease [262].
*Puccinia triticina-*Wheat	Change analysis in the proteomes of both host and pathogen during development of wheat leaf rust disease [154].
*Rhizoctonia solani-*Rice	Identification of proteins and DNA markers in rice associated with response to infection by *R. solani* [263].
Rust-Phaseolus vulgaris	Study of basal and R-gene-mediated plant defense in bean leaves against this pathogen [264].
S. sclerotiorum-Brassica napus	Study of changes in the leaf proteome of *B. napus* accompanying infection by *S. sclerotiorum* [265].

(for reviews see [266–268]), although the plant-fungus association has been the one most studied by Proteomics approaches (Table 6), which is outside the scope of this review. On the other hand, some fungal species have attracted an increasing interest in the biotechnological industry, in food science, or in agronomy as biocontrol agents (Table 7), which is also beyond the objectives of this work. At this point, this review describes studies published up to December 2009 in plant pathogenic fungi in descriptive proteomics (intracellular proteomics, subproteomics, and secretomics), differential expression proteomics, as well as some basic knowledge about the Interactomics in fungi (Table 3).

3.2.1. Descriptive, Subcellular, and Differential Expression Proteomics. Within this section, papers devoted to establishing reference proteome maps of fungal cells and structures and subcellular fractions, and to study changes in the protein profile between species, races, populations, mutants, growth and developmental stages, as well as growth conditions, are discussed, paying special attention to proteins related to pathogenicity and virulence.

Most of the reported work mainly uses mycelia from in vitro grown fungi, and 2-DE coupled to MS as proteomic strategy. Thus, a partial proteome map has been reported for the ascomycete *B. cinerea* Pers. Fr. (teleomorf *Botryotinia fuckeliana* (de Bary) Whetzel), a phytopathogenic

TABLE 7: Original proteomics papers published on fungi for biotechnological or agricultural applications.

Fungus	Interest (reference)
Boletus edilus	Study of salinity stress of this ectomycorrhizal fungus for its importance in reforestation in saline areas [269].
Coprinopsis cinerea	
Pleorotus ostreatus	
Phanerochaete chrysosporium	Optimization of a protocol for 2-DE of extracellular proteins from these wood-degrading fungi [147].
Polyporus brumalis	
Schizophyllum commune	
Glomus intraradices	Study of arbuscular mycorrhiza symbiosis [173].
Metarhizium anisopliae	Study of bioinsecticidal activity of this fungus to develop novel compounds or produce genetically modified plants resistant to insect pests [270].
Monascus pilosus	Study of the influence of nitrogen limitation for industrial production of many poliketide secondary metabolites [271].
Phanerochaete chrysosporium	Several studies of ligninolytic processes for wood biodelignification in cellulose pulp industries [130, 131, 149, 150, 272].
Pleurotus sapidus	Study of secretome for wood biodelignification for peanuts industry applications [152].
Trichoderma atroviride	
T. harzianum	Several studies in these fungus for their biocontrol properties [126, 273–275]
T. reesei	Study of cell-wall-degrading enzymes in *A. bisporigera* (an ectomycorrhizal basidiomycetous fungus) comparing with a MS/MS-based shotgun proteomics of the secretome of *T. reesei* [276].
Amanita bisporigera	

necrotroph pathogen causing significant yield losses in a number of crops, by Fernández-Acero et al. They have reported the detection of 400 spots in Coomassie-stained 2-DE gels, covering the 5.4–7.7 pH and 14–85 kDa ranges. Out of 60 spots subjected to MS analysis, twenty-two proteins were identified by MALDI-TOF or ESI IT MS/MS, with some of them corresponding to forms of malate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, and a cyclophilin, proteins that have been related to virulence [45]. In a second study, comparative proteomic analysis of two *B. cinerea* strains differing in virulence and toxin production revealed the existence of qualitative and quatitative differences in the 2-DE protein profile. Some of them were the same proteins mentioned above and they appeared overexpressed or exclusively in the most virulent strain [46]. A third and more exhaustive work tried to establish a proteomic map of *B. cinerea* during cellulose degradation [47]. Using 2-DE and MALDI-TOF/TOF MS/MS, 306 proteins were identified, mostly representing unannotated proteins. The authors conclude that since cellulose is one of the major components of the plant cell wall, many of the identified proteins may have a crucial role in the pathogenicity process, be involved in the infection cycle, and be potential antifungal targets.

A close relative to *B. cinerea* is the soil-borne *Sclerotinia sclerotiorum*. Yajima and Kav [64] performed the first comprehensive proteome-level study in this important phytopathogenic fungus, in order to gain a better understanding of its life cycle and its ability to infect susceptible plants. For the high-throughput identification of secreted as well as mycelial proteins, they employed 2-DE and MS/MS. Eighteen secreted and 95 mycelial proteins were identified. Many of the annotated secreted proteins were cell wall-degrading

enzymes that had been previously identified as pathogenicity or virulence factors of *S. Sclerotiorum*. Furthermore, this study has allowed the annotation of a number of proteins that were unnamed, predicted, or hypothetical proteins with undetermined functions in the available databases.

Xu et al. [50] analyzed the proteome profile of six different isolates of *Curvularia lunata*, a maize phytopathogenic fungus, by means of 1-DE and 2-DE, in an attempt to correlate the band or spot pattern with virulence. According to the 1-DE band pattern, isolates were clustered into three groups consisting of different virulent types. By 2-DE 423 spots were resolved with 29 of them being isolate-specific, and 39 showed quantitative differences. Twenty proteins were identified by MALDI-TOF-TOF, most of them associated with virulence differentiation, metabolisms, stress response, and signal transduction. One of them was identified as Brn1 protein which has been reported to be related to melanin biosynthesis and the virulence differentiation in fungi.

The fungal pathogen *F. graminearum* (teleomorph *Gibberella zeae*) is the causal agent of *Fusarium* head blight in wheat, barley, and oats and *Gibberella* ear rot in maize in temperate climates worldwide. It synthesizes trichothecene mycotoxins during plant host attack to facilitate spread within the host. In order to study proteins and pathways that are important for successful host invasion, Taylor et al. [53] conducted experiments in which *F. graminearum* cells were grown in aseptic liquid culture conditions conducive to trichothecene and butenolide production in the absence of host plant tissue. Protein samples were extracted from three biological replicates of a time course study and subjected to iTRAQ (isobaric tags for relative and absolute quantification) analysis. Statistical analysis of a filtered dataset of 435 proteins revealed 130 *F. graminearum* proteins that

exhibited significant changes in expression, 72 of which were upaccumulated relative to their level at the initial phase of the time course. There was good agreement between upaccumulated proteins identified by 2-DE-MS/MS and iTRAQ. RT-PCR and northern hybridization confirmed that genes encoding proteins that were upregulated based on iTRAQ were also transcriptionally active under mycotoxin-producing conditions. Numerous candidate pathogenicity proteins were identified using this technique, including many predicted secreted proteins. Curiously, enzymes catalyzing reactions in the mevalonate pathway leading to trichothecene precursors were either not identified or only identified in one replicate, indicating that proteomics approaches cannot always probe biological characteristics. Two-DE with MS has been used to compare the proteome of virus-free and virus-(FgV-DK21-) infected *F. graminearum* cultures [54]. The virus perturbs fungal developmental processes such as sporulation, morphology, and pigmentation and attenuates its virulence. A total of 148 spots showing differences in abundance were identified. Among these spots, 33 spots were subjected to ESI-MS/MS, with 23 identified. Seven proteins including sporulation-specific gene SPS2, triose phosphate isomerase, nucleoside diphosphate kinase, and woronin body major protein precursor were upaccumulated while 16, including enolase, saccharopine dehydrogenase, flavohemoglobin, mannitol dehydrogenase, and malate dehydrogenase, were downaccumulated. Variations in protein abundance were investigated at the mRNA level by real-time RT-PCR analysis, which confirmed the proteomic data for 9 out of the representative 11 selected proteins.

There are a few proteomics studies on fungal spores published. The results in the characterization of *Penicillium* spores by MALDI-TOF MS with different matrices demonstrated its ability for the classification of fungal spores [277]. Recently, Sulc et al. [40] have reported protein profiling of intact *Aspergillus* ssp. spores, including some plant pathogenic species, by MALDI-TOF MS, and they built up a mass spectral database with twenty-four *Aspergillus* strains. Thus, these mass finger-printing generated by MS can be used for typing and characterizing different fungal strains and finding new biomarkers in host-pathogen interactions. Another study on *B. graminis* f.sp. *hordei* (*Bgh*) spores concluded with the first proteome of *Bgh*, using a combination of 2-DE and MS analyses and matched to NCBInr EST on *Bgh*-translated genome databases [43]. The identity of 123 distinct fungal gene products was determined, most of them with a predicted function in carbohydrate, lipid, or protein metabolism indicating that the conidiospore is geared for the breakdown of storage compounds and protein metabolites during germination correlating with previously reported transcriptomic data [278, 279]. These results allowed a functionally annotated reference proteome for *Bgh* conidia.

Holzmuller et al. [118] have reported a technique to isolate the fungal haustorium (specialised structures, existing in intimate contact with the host cell, are required by the pathogen to acquire nutrients from the host cell) from infected plants, using the barley powdery mildew as an experimental system. The technique is of relevance in the study of the molecular bases of biotrophy considering that biotrophic fungi, including downy mildews (Oomycota), powdery mildews (Ascomycota), and rust fungi (Basidiomycota), are some of the most destructive pathogens on many plants. Extracted proteins were separated and analyzed by LC-MS/MS. The searches were made against a custom *Bgh* EST sequence database and the NCBInr fungal protein database, using the MS/MS data, and 204 haustorium proteins were identified. The majority of the proteins appeared to have roles in protein metabolic pathways and biological energy production. Surprisingly, pyruvate decarboxylase (PDC), involved in alcoholic fermentation and commonly abundant in fungi and plants, was absent in both their *Bgh* proteome data set and in their EST sequence database. Significantly, BLAST searches of the recently available *Bgh* genome sequence data also failed to identify a sequence encoding this enzyme, strongly indicating that *Bgh* does not have a gene for PDC [44].

In order to overcome the low proteome coverage of most of the proteomic platforms available, this being related to the physicochemical and biological complexity and high dynamism range of proteins, different strategies directed at subfractionating the whole proteome have been developed, most of them involving cell fractionation. The analysis of the subcellular proteomes [280] not only allows a deeper proteome coverage but also provides relevant information on the biology of the different organules, protein location, and trafficking. The number of intracellular subproteomic studies carried out with fungal plant pathogens is minimum. Next we introduce a couple of papers appearing in the literature. The number of them devoted to the cell wall and extracellular fraction is much higher, and because of that a specific section is devoted to them. Hernández-Macedo et al. [131] have reported differences in the patterns of cellular and membrane proteins obtained from iron-sufficient and iron-deficient mycelia from *P. chrysosporium* and *L. edodes* by using SDS-PAGE and 2-DE. Mitochondria have also received attention. Grinyer et al. [121] were the first to publish a mitochondrial subproteome, describing a successful sample preparation protocol and mitochondrial proteome map for *T. harzianum*. Based on protein databases of *N. crassa*, *A. nidulans*, *A. oryzae*, *S. cerevisiae*, and *Schizosaccharomyces pombe*, they identified 25 unique mitochondrial proteins involved in the tricarboxylic acid cycle, chaperones, binding-proteins and transport proteins, as well as mitochondrial integral membrane proteins. More recently, the same researchers separated and identified 13 of the 14 subunits of the *T. reesei* 20S proteasome [281], providing the first filamentous fungal proteasome proteomics and paving the way for future differential display studies addressing intracellular degradation of endogenous and foreign proteins in filamentous fungi.

Relevant information on biological systems and processes comes from comparative studies in which genotypes, including mutants, developmental stages, or environmental conditions supply the knowledge inferred from the observed differences. Fungal pathogenicity requires the coordinated regulation of multiple genes (and their protein products) involved in host recognition, spore germination, hyphal

penetration, appressorium formation, toxin production, and secretion. To study the infection cycle and to identify virulence factors, proteomics provides us with a powerful tool for analyzing changes in protein expression between races and stages. However, most of these studies are made *in planta* after the plant inoculation, which is outside the scope of this review. In the case of plant fungal pathogens at least four papers have reported changes in the proteome at different developmental stages or strains. The dimorphic phytopathogenic fungus *Ustilago maydis* has been established as a valuable model system to study fungal dimorphism and pathogenicity. In its haploid stage, the fungus is unicellular and multiplies vegetatively by budding and undergoing a dimorphic transition infective filamentous growth. This process is coordinately regulated by the bW/bE transcription factor. Böhmer et al. [68] reported the first proteome reference map of *U. maydis* cells, in which proteins were identified combining 2-DE with MALDI-TOF MS and ESI-MS/MS analyses. The authors observed 13 proteins spots accumulated in greater abundance in the bW/bE-induced filamentous form than in the budding state. The majority of the identified proteins might have putative roles in energy and general metabolism. Comparison of Rac1- and -b-regulated protein sets supports the hypothesis that filament formation during pathogenic development occurs *via* stimulation of a Rac1-containing signalling module. The proteins identified in this study might prove to be potential targets for antibiotic substances specifically targeted at dimorphic fungal pathogens. Detailed information on the proteins can be found in an interactive map accessible at the MIPS Ustilago Maydis Database (MUMDB; http://mips.gsf.de/genre/proj/ustilago/Maps/2D/). The reference map generated from *U. maydis* had a coverage of 4% of all annotated genes, indicating the low proteome coverage encompassed by standard proteomic techniques.

By using 2-DE and MALDI-TOF MS, specific proteins of asexual life stages from *Phytophthora palmivora*, a pathogen of cocoa and other economically important tropical crops, were analyzed [58]. From 400 (cyst and germinated cyst) to 800 (sporangial) could be resolved. Approximately 1% of proteins appeared to be specific for each of the mycelial, sporangial, zoospore, cyst, and germinated cyst stages of the life cycle. Moreover, they made the protein profiles of parallel samples of *P. palmivora* and *P. infestans* and demonstrated that precisely 30% of proteins comigrated suggesting that proteomics could be used to proteptype *Phytophthora* spp. In this work, only three identified proteins were reported, corresponding to actin isoforms. More recently, Ebstrup et al. [59] performed a proteomic study of proteins from cysts, germinated cysts, and appressoria on *P. infestans* grown in vitro, identifying significant changes in the amount of several proteins. These identified proteins were most likely important for disease establishment and some of the proteins could therefore be putative targets for disease control. For example, downregulation of the crinkling- and necrosis-inducing (CRN2) protein in appressoria compared to germinated cysts and the discovery of upregulation of a putative elongation factor (EF-3) are of great interest. On the one hand, CRN2 protein might have an important function in the interaction

with the host-plant before and after penetration into the leaf, this being a putative target for disease control. Since plants presumably do not contain EF-3, it could represent a putative antioomycete as well as a putative antifungal target. Furthermore, several representatives of housekeeping systems were upaccumulated, and these changes are most likely involved in the runup to the establishment of the infection of the host plant. The biotrophic fungal pathogen *U. appendiculatus* is the causal agent of rust disease of beans. Cooper et al. [67] surveyed the proteome from germinating and ungerminated asexual uredospores of this pathogen, using MudPIT MS/MS. The proteins identified revealed that uredospores require high energy and structural proteins during germination, indicating a metabolic transition from dormancy to germination.

The role of signal transduction in the pathogenecity of *Stagonospora nodorum* is well established and the inactivation of heterotrimeric G protein signaling caused developmental defects and reduced pathogenicity [282]. In a followup study, the *S. nodorum* wild-type and Galpha-defective mutant (gna1) proteomes were compared via 2-DE coupled to LC-MS/MS. By matching the protein mass spectra to the translate *S. nodorum* genome, the study identified several Gna1-regulated proteins, including a positively regulated short-chain dehydrogenase (Sch1) [65].

Cao et al. [61] released an evaluation of pathogenic ability of *Pyrenophora tritici-repentis* and the possible adaptation to a saprophytic habit of an avirulent race. This fungus causes tan spot, an important foliar disease of wheat, and produces multiple host-specific toxins, including Ptr ToxB, which is also found in avirulent isolates of the fungus. In order to improve the understanding of the role of this homolog and evaluate the general pathogenic ability of *P. tritici-repentis*, the authors compared both full mycelial and secreted proteomes of avirulent and virulent isolates of the pathogen, by 2-DE and ESI-q-TOF MS/MS. The proteomic analysis revealed a number of the proteins found to be upregulated in a virulent race, which has been implicated in microbial virulence in other pathosystems, such as the secreted enzymes a-mannosidase and exo-b-1,3-glucanase, heat-shock and bip proteins, and various metabolic enzymes, which suggests a reduced general pathogenic ability in avirulent race of *P. tritici-repentis*, irrespective of toxin production.

3.2.2. Extracellular and Cell Wall Proteins: The Secretome. Most eukaryotic plant pathogens initially invade the space between host cell walls (the apoplastic space), and much of the initial host defence and pathogen counter defence happens in the apoplast and commonly involves secreted pathogen and host-derived proteins and metabolites [283]. While some pathogens remain exclusively in the apoplast, such as *Cladosporium fulvum*, others, including mildews, rusts smuts, *Phytophthora*, and *Magnaporthe* species, breach host cell walls but remain external to and separated from the host cytoplasm by host and pathogen cell membranes. Some host wall-breaching pathogens, like rusts, mildews, and oomycetes, form specialised expanded hyphal protuberances

called haustoria whereas others, like maize smut and the rice blast fungi, use unexpanded but probably specialised intrahost cell wall hyphae [284]. The role of these structures was initially thought to be primarily nutrient acquisition, but recently their additional role in secretion of effectors, some of which are translocated to the host cytoplasm, has become more apparent. These issues have been recently reviewed by Ellis et al. [113].

The secretome has been defined as being the combination of native secreted proteins and the cell machinery involved in their secretion [285]. A defining characteristic of plant pathogenic fungi is the secretion of a large number of degradative enzymes and other proteins, which have diverse functions in nutrient acquisition, substrate colonization, and ecological interactions [286–288]. Several extracellular fungal enzymes, such as polygalacturonase, pectate lyase, xylanase, and lipase, have been shown or postulated to be required for virulence in at least one host-pathogen interaction [289–295]. Proteomics is the right approach to study the interaction between plants and microbes mediated by excreted molecules, the role of the cell wall and the interface, and to identify fungal protein effectors facilitating either infection (virulence factors, enzymes of the toxin biosynthesis pathways) or trigger defence responses (avirulence factors). In the light of this, it has been said that, unlike animals, "fungi digest their food and [then] eat it" [296], illustrating the large number of extracellular hydrolytic enzymes necessary to digest a plethora of potential substrates. Therefore, many of these proteins are of special interest in the study of plant pathogens [46, 64]. This might also be owing to the fact that secretome sample preparation is much faster and simpler than extraction and preparation of intracellular proteins. Next, a number of papers covering this topic are presented, including those dealing with the secretome of *Trichoderma* spp, a study directed at identifying proteins related to its biofungicidal activity.

Pioneering work on this field comes before the arrival of proteomics during the 1990s, with typical studies focused on the identification, purification, and characterization of single secreted proteins, under the influence of the biotechnology industry for the production of enzymes for commercial and industrial use [297]. The first complete proteomic study of secreted proteins was released on the filamentous fungus *A. flavus* [41, 42]. The interest of this study was the ability of both *A. flavus* and *A. parasiticus* to degrade the flavonoids that plants produce as typical secondary metabolites against invading microorganisms. The secreted proteins were analyzed by 2-DE and MALDI-TOF mass spectrometry, with 15 rutin-induced proteins and 7 noninduced proteins identified, among them enzymes of routine catabolism pathway and glycosidases.

In *F. graminearum*, a devastating pathogen of wheat, maize, and other cereals, Phalip et al. [51] investigated the exoproteome of this fungus grown on glucose and on plant cell wall (*Humulus lupulus*, L.). The culture medium was found to contain a larger amount of proteins and these were more diverse when the fungus grew on the cell wall. Using both 1-DE and 2-DE coupled to LC-MS/MS analysis and protein identification based on similarity searches,

84 unique proteins were identified in the cell wall-grown fungal exoproteome and 45% were implicated in plant cell wall degradation. These cell wall-degradating enzymes were predominantly matches to putative carbohydrate active enzymes implicated in cellulose, hemicelluloses, and pectin, catabolism. As expected, *F. graminearum* grown on glucose produced relatively few cell wall-degrading enzymes. These results indicated that fungal metabolism becomes oriented towards the synthesis and secretion of a whole arsenal of enzymes able to digest almost the complete plant cell wall.

The secretome has also been analyzed in *S. sclerotium* as commented above [64]. In this study, 52 secreted proteins were identified and many of the annotated secreted proteins were cell wall-degrading enzymes that had been identified previously as pathogenic or virulence factors of *S. sclerotium*. However, one of them, α-L-arabinofuranosidase, which is involved in the initiation or progression of plant diseases, was not detected by previous EST studies, clearly demonstrating the merit of performing proteomic research.

Two studies have been published reporting the *B. cinerea* (B05.10) secreted proteins analysis [48, 49]. First, secretions were collected from fungus grown on a solid substrate of cellophane membrane while mock infecting media supplemented with the extract of full red tomato, ripened strawberry, or *Arabidopsis* leaf extract. Overall, 89 *B. cinerea* proteins were identified by high-throughput LC-MS/MS from all growth conditions. Sixty of these proteins were predicted to contain a SignalP motif indicating the extracellular location of the proteins. The proteins identified were transport proteins, proteins well-characterized for carbohydrate metabolism, peptidases, oxidation/reduction, and pathogenicity factors that could provide important insights into how *B. cinerea* might use secreted proteins for plant infection and colonization [48]. In the second work, the impact of degree of esterification of pectin on secreted enzyme of *B. cinerea* was studied, because changes during the ripening process of fruits appear to play an important role in the activation of the dormant infection. All the major components of the fruit cell wall (pectin, cellulose, hemicellulose) undergo these changes. By 1-DE and LC-MS/MS, 126 proteins were identified and 87 proteins were predicted secreted by SignalP, some of them being pectinases. The results showed that the growth of *B. cinerea* and the secretion of proteins were similar in cultures containing differently esterified pectins, and therefore it is likely that the activation of this fungi from dormant state is not solely dependent on changes in the degree of esterification of the pectin component of the plant cell wall [49]. Therefore, future studies of the *B. cinerea* secretome in infections of ripe and unripe fruits will provide important information for describing the mechanisms that the fungus employs to access nutrients and decompose tissues.

Using the plant pathogenic fungus *L. maculans* and symbiont *Laccaria bicolor* grown in culture, Vincent et al. [56] established a proteomic protocol for extraction, concentration, and resolution of the fungal secretome. These authors used both broad and narrow acidic and basic pH range in IEF. The quality of protein extracts was assessed by both 1-DE and 2-DE and MS identification. Compared

with the previously published protocols for which only dozens of 2-DE spots were recovered from fungal secretome samples, in this study, up to approximately 2000 2-DE spots were resolved. This high resolution was confirmed with the identification of proteins along several pH gradients as well as the presence of major secretome markers such as endopolygalacturonases, beta-glucanosyltransferases, pectate lyases, and endoglucanases. Thus, shotgun proteomic experiments evidenced the enrichment of secreted protein within the liquid medium.

One of the earliest works was released on *Trichoderma reesei* mycelium cell wall, one of the most powerful producers of extracellular proteins, this study being justified in order to find out the protein secretory pathways and the effect of the fungal genus, strain, and media condition on the excretion through the cell wall [298]. A total of 220 cell envelope-associated proteins were successfully extracted and separated by 2-DE from *Trichoderma reesei* mycelia actively secreting proteins and from mycelia in which the secretion of proteins is low. Out of the 52 2-DE spots subjected to ESI-TOF MS, 20 were identified, with HEX1, the major protein in Woronin body, a structure unique to filamentous fungi, being the most abundant one. Suárez et al. [299] studied the secretome of *T. harzianum* grown using either chitin (a key cell wall component) or cell wall of other fungi (*R. solani*, *B. cinerea*, or *Pythium ultimum*) as a nutrient source. For each different substrate, they found significant differences in 2-DE maps of extracellular proteins. However, despite these differences, the most abundant protein under all conditions was a novel aspartic protease (P6281), which showed a strong homology with polyporopepsin from *Irpex lacteus*. This led to speculation that this protein plays a fundamental role in the parasitic activity of *Trichoderma* spp. Marra et al. [273] have studied interactions between *T. atroviride*, two different fungal phytopathogens (*B. cinerea* and *R. solani*), and plants (bean). Two-DE was used to analyze separately collected proteomes from each single, two- or three-partner interaction. Then, differential proteins were subjected to MALDI-TOF MS and in silico analysis to search homologies with known proteins. Thus, a large number of protein factors associated with the multiplayer interactions examined were identified, including protein kinases, cyclophilines, chitine synthase, and ABC transporters. Recently, another similar study was released between *T. harzianum* and *R. solani* by analysing the secretome to identify the target proteins that are directly related to biocontrol mechanism [274]. Seven cell-wall degrading enzymes, chitinase, cellulase, xylanase, β-1,3-glucanase, β-1,6-glucanase, mannanase, and protease, were revealed by activity assay, in-gel activity stain, 2-DE, and LC-MS/MS analysis, these being increased in response to *R. solani*.

A cell wall proteome has been proposed for the oomycete *Phytophthora ramorum*, the causal agent of sudden oak death, in order to study its pathogenic factors [60]. This study showed an inventory of cell wall-associated proteins based on MS sequence analysis. Seventeen secreted proteins were identified by homology searches. The functional classification revealed several cell wall-associated proteins, thus suggesting that cell wall proteins may also be important for fungal pathogenicity.

The filamentous fungus *Neurospora crassa* is a model laboratory organism but in nature is commonly found growing on dead plant material, particularly grasses. Using functional genomics resources available for *N. crassa*, which include a near-full genome deletion strain set and whole genome microarrays, Tian et al. [300] undertook a system-wide analysis of plant cell wall and cellulose degradation, identifying approximately 770 genes that showed expression differences when *N. crassa* was cultured on ground *Miscanthus* stems as a sole carbon source. An overlap set of 114 genes was identified from expression analysis of *N. crassa* grown on pure cellulose. Functional annotation of upregulated genes showed enrichment for proteins predicted to be involved in plant cell wall degradation, but also many genes encoding proteins of an unknown function. As a complement to expression data, the secretome associated with *N. crassa* growth on Miscanthus and cellulose was determined using a shotgun MudPIT proteomic strategy. Over 50 proteins were identified, including 10 of the 23 predicted *N. crassa* cellulases. Strains containing deletions in genes encoding 16 proteins detected in both the microarray and mass spectrometry experiments were analyzed for phenotypic changes during growth on crystalline cellulose and for cellulase activity. While growth of some of the deletion strains on cellulose was severely diminished, other deletion strains produced higher levels of extracellular proteins that showed increased cellulase activity. These results show that proteomics in combination with other powerful tools available in model systems such as *N. crassa* allow for a comprehensive system level understanding of fungal biology.

3.2.3. Interactomics. The biological organization in living cells can be regarded as being part of a complex network [301–303]. Traditional approaches studied a single gene or unique protein and therefore did not provide a complete knowledge of the biological processes. Proteins release their functional roles through their interactions with one another in vivo. Thus, developing a protein-protein interaction (PPI) network can lead to a more comprehensive understanding of the cell processes [304]. Interactomics is a discipline at the intersection of bioinformatics and biology that deals with studying both the interactions and the consequences of those interactions between and among proteins and other molecules within a cell [305]. The network of all such interactions is called the interactome. Interactomics thus aims to compare these interaction networks (i.e., interactomes) between and within species in order to find how the traits of such networks are either preserved or varied. Interactomics is an example of top-down systems biology, which takes an overhead, as well as overall, view of a biosystem or organism.

In recent years, high-throughput methods have been implemented to identify PPIs [306–310] and these have recently been reviewed in [311, 312]. Using these experimental methods, such as yeast two-hybrid screens, PPI networks for a series of model organisms were determined and allowed

us to better understand the function of proteins at the level of system biology. Two-hybrid screening (also known as yeast two hybrid system or Y2H) is a powerful tool for identifying PPI. The premise behind the test is the activation of dowstream reporter gene(s) by the binding of a transcription factor onto an upstream activating sequence (UAS). For the purposes of two-hybrid screening, the transcription factor is split into two separate fragments, called the binding domain (BD) and activating domain (AD). The BD is the domain responsible for binding to the UAS and the AD is the domain responsible for activation of transcription [313].

In parallel with the large-scale experimental determination of PPI, many PPI prediction methods were also developed. These methods are based on diverse attributes, concepts, or data types, such as interolog [314], gene expression profiles [315], gene ontology (GO) annotations [316], domain interactions [317], coevolution [318], and structural information [319]. Some machine learning methods, such as support vector machines (SVMs) have also been used to predict PPIs [320, 321]. Among the above-mentioned computational methods the interolog approach has been widely implemented [322] and has proved to be reliable for predicting PPI from model organisms [323]. The core idea of the interolog approach is that many PPIs are conserved in different organisms [324]. Accumulated PPI data from model organisms as well as advances in detecting orthologous proteins in different organisms [280] have continuously made the interolog method an increasingly powerful tool for constructing PPI maps for entire proteomes.

Using the interolog method, He et al. [57] constructed the first PPI network for *M. grisea*. Thus, 11674 PPIs among 3017 *M. grisea* proteins were deduced from the experimental PPI data in different organisms, although the predicted PPI network covered approximately only one-fourth of the fungal proteome and may still contain many false-positives. Moreover, they built two subnets called pathogenicity and secreted proteins networks, which may be helpful in constructing an interactome between the rice blast fungus and rice (MPID website, http://bioinformatics .cau.edu.cn/zzd_lab/MPID.html).

A *F. graminearum* protein-protein interaction database providing comprehensive information on protein-protein interactions based on both interologs from several protein-protein interaction of seven species and domain-domain interactions experimentally determined based on protein is available at http://csb.shu.edu.cn/fppi [55]. It contains 223 166 interactions among 7406 proteins for *F. graminearum*, covering 52% of the whole *F. graminearum* proteome.

4. Concluding Remarks

In the current scientific scenario, Proteomics should be understood to be part of a multidisciplinary approach. A combination of high-throughput "Omics" (Genomics, Transcriptomics, Proteomics, and Metabolomics) and classical biochemistry and cell biology techniques should be used for data validation and to deepen the knowledge of living organisms. Proteomic techniques are used to characterize

a specific protein or a structural or functional group of proteins. This is what we can call "Hypothesis-driven Proteomics", "Targeted Proteomics", or "Proteinomics". This type of study will provide relevant information on protein structure and function, isoforms, organs, cells, and subcellular location and trafficking, processing, signal peptides, PTMs, expression kinetics, and correlation with RNA and metabolites. At the same time, it is a method for validating data obtained using one specific approach [102].

Despite the continuous development and improvement of powerful proteomic techniques, protocols, equipments, and bioinformatic tools, just a minimal fraction of the cell proteome, and for only a few organisms, has been characterized so far. This is mainly due to the enormous diversity and complexity of proteomes, and to technical limitations in quantification, sensitivity, resolution, speed of data capture, and analysis.

In the field of Fungal MS-based Proteomics, great progress has been made in past years. This is because of the increasing number of fungal genomes available and the developments in sample preparation, high-resolution protein separation techniques, MS, MS software for effective protein identification and characterization, and bioinformatics technology. The tremendous diversity and genome flexibility in fungi, however, will make this task a difficult one. Thus, a key step in sequence analysis is the annotation. The existing programs for automated gene prediction are not perfect and need to be improved or trained better. Follow-up manual annotation is also necessary to improve the accuracy of automated annotation, but this is time-consuming and labor-intensive. Ultimately, a comprehensive genome database similar to YPD (http://www.yeastgenome.org/) will be desirable for fungal pathogens.

To date, most proteomic studies in plant pathogenic fungi have been limited to 1- and 2-DE analysis. However, various powerful proteomic methods have been developed for genomewide analysis of protein expression, protein localization, and protein-protein interaction in fungi. Whole-genome protein arrays and systematic yeast two-hybrid assays have been used to characterize the yeast proteome and interactome. Integration of large-scale genomics and proteomics data enables the elucidation of global networks and system biology studies in yeast. It is necessary for similar advanced proteomic resources to be soon available for some fungal plant pathogens. Moreover, second-generation MS technologies for Quantitative Proteomics such as 2-DIGE, stable isotope labeling, (ICAT, iTRAQ and SILAC), or label-free methods (peak integration, spectral counting) have not yet begun or are beginning to be applied to Fungal Proteomics research.

Otherwise, the major challenge is the analysis and significance of PTMs because proteins have properties arising from their folded structure and so generic methods are difficult to design and apply.

In conclusion, since plant pathogenic fungi cause important losses in a number of crops, it is necessary to make high-throughput studies on these organisms to identify pathogenicity factors. Although genomics-based investigation of host-pathogen interactions can provide valuable

information on the changes in gene expression, the investigation into changes in protein abundance is also important, in order to identify those proteins that are essential during such interactions. This is because there is often a poor correlation between transcript and protein abundance [325]. Proteomics analysis is an excellent tool that can give us a great deal of information about fungal pathogenicity by high-throughput studies. This approach has allowed the identification of new fungal virulence factors, characterizing signal transduction or biochemical pathways, studying the fungal life cycle and their life-style. We can use this information to provide new targets for disease crop diagnosis focused on fungicide design. Otherwise, the secretome analysis is especially important because fungi secrete an arsenal of extracellular enzymes to break down the plant cell wall for pathogen penetration and nutrient consumption. In this sense, Proteomics allows us to identify numerous differential proteins involved in multiple-player cross-talk normally occurring in nature between plants and pathogens, the so-called "interaction proteomes". Finally, MS-based Proteomics can help us to characterize fungal strains and find new biomarkers in host-pathogen interactions.

In short, Fungal Proteomics is in the first step. Therefore, we still have a long way to go in the "Omics" of Plant Pathogenic Fungi compared to studies made in Humans, Bacteria, Yeast, or Plants. The important investment made by both the public and private sector in recent years augurs good prospects in fungal proteomics research in the future.

Abbreviations

1-DE:	One-dimentional electrophoresis
2-DE:	Two-dimentional electrophoresis
ESI:	Electrospray ionization
HPLC:	High performance liquid chromatograohy
IEF:	Isoelectrofocusing
LC:	Liquid chromatography
MALDI:	Matrix-assisted laser desorption/ionization
MS:	Mass spectrometry
SDS-PAGE:	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TOF:	Time of flight.

Acknowledgments

References

[1] D. L. Hawksworth, "The fungal dimension of biodiversity: magnitude, significance, and conservation," *Mycological Research*, vol. 95, no. 6, pp. 641–655, 1991.

[2] D. L. Hawksworth, P. M. Kirk, B. C. Sutton, and D. N. Pegler, *Ainsworth and Bisby's Dictionary of the Fungi*, International Mycological Institute, Surrey, UK, 2001.

[3] G. N. Agrios, *Plant Pathology*, Elsevier Academic Press, San Diego, Calif, USA, 2005.

[4] A. M. Green, U. G. Mueller, and R. M. Adams, "Extensive exchange of fungal cultivars between sympatric species of fungus-growing ants," *Molecular Ecology*, vol. 11, no. 2, pp. 191–195, 2002.

[5] A. Pérez-García, E. Mingorance, M. E. Rivera, et al., "Long-term preservation of *Podosphaera fusca* using silica gel," *Journal of Phytopathology*, vol. 154, no. 3, pp. 190–192, 2006.

[6] D. K. Arora, P. D. Bridge, and D. Bhatnagar, *Fungal Biotechnology in Agricultural, Food and Environmental Applications*, Marcel Dekker, New York, NY, USA, 2004.

[7] D. A. Glawe, "The powdery mildews: a review of the world's most familiar (yet poorly known) plant pathogens," *Annual Review of Phytopathology*, vol. 46, pp. 27–51, 2008.

[8] S. Isaac, *Fungal-Plant Interactions*, Chapman & Hall, London, UK, 1992.

[9] H. H. Flor, "Inheritance of pathogenicity in *Melampsora lini*," *Phytopathology*, vol. 32, pp. 653–669, 1942.

[10] P. J. De Wit, R. Mehrabi, H. A. Van den Burg, and I. Stergiopoulos, "Fungal effector proteins: past, present and future," *Molecular Plant Pathology*, vol. 10, no. 6, pp. 735–747, 2009.

[11] H. P. van Esse, J. W. Van't Klooster, M. D. Bolton, et al., "The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense," *Plant Cell*, vol. 20, no. 7, pp. 1948–1963, 2008.

[12] T. L. Friesen, J. D. Faris, P. S. Solomon, and R. P. Oliver, "Host-specific toxins: effectors of necrotrophic pathogenicity," *Cellular Microbiology*, vol. 10, no. 7, pp. 1421–1428, 2008.

[13] X. Gao and M. V. Kolomiets, "Host-derived lipids and oxylipins are crucial signals in modulating mycotoxin production by fungi," *Toxin Reviews*, vol. 28, no. 2-3, pp. 79–88, 2009.

[14] K. S. Kim, J. Y. Min, and M. B. Dickman, "Oxalic acid is an elicitor of plant programmed cell death during *Sclerotinia sclerotiorum* disease development," *Molecular Plant-Microbe Interactions*, vol. 21, no. 5, pp. 605–612, 2008.

[15] C. B. Lawrence, T. K. Mitchell, K. D. Craven, Y. Cho, R. A. Cramer, and K. H. Kim, "At death's door: alternaria pathogenicity mechanisms," *Plant Pathology Journal*, vol. 24, no. 2, pp. 101–111, 2008.

[16] H. C. van der Does and M. Rep, "Virulence genes and the evolution of host specificity in plant-pathogenic fungi," *Molecular Plant-Microbe Interactions*, vol. 20, no. 10, pp. 1175–1182, 2007.

[17] Z. Zhang, T. L. Friesen, K. J. Simons, S. S. Xu, and J. D. Faris, "Development, identification, and validation of markers for marker-assisted selection against the Stagonospora nodorum toxin sensitivity genes Tsn1 and Snn2 in wheat," *Molecular Breeding*, vol. 23, no. 1, pp. 35–49, 2009.

[18] B. Pariaud, V. Ravigné, F. Halkett, H. Goyeau, J. Carlier, and C. Lannou, "Aggressiveness and its role in the adaptation of plant pathogens," *Plant Pathology*, vol. 58, no. 3, pp. 409–424, 2009.

[19] D. Parker, M. Beckmann, H. Zubair, et al., "Metabolomic analysis reveals a common pattern of metabolic reprogramming during invasion of three host plant species by *Magnaporthe grisea*," *Plant Journal*, vol. 59, no. 5, pp. 723–737, 2009.

[20] S. McCouch, "Diversifying selection in plant breeding," *PLoS Biology*, vol. 2, no. 10, article e347, 2004.

[21] Y. Yang, H. Zhang, G. Li, W. Li, X. Wang, and F. Song, "Ectopic expression of MgSM1, a Cerato-platanin family protein from *Magnaporthe grisea*, confers broad-spectrum disease resistance in *Arabidopsis*," *Plant Biotechnology Journal*, vol. 7, no. 8, pp. 763–777, 2009.

[22] B. S. Kim and B. K. Hwang, "Microbial fungicides in the control of plant diseases," *Journal of Phytopathology*, vol. 155, no. 11-12, pp. 641–653, 2007.

[23] G. Berg, "Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture," *Applied Microbiology and Biotechnology*, vol. 84, no. 1, pp. 11–18, 2009.

[24] G. Aguileta, M. E. Hood, G. Refrégier, and T. Giraud, "Chapter 3 genome evolution in plant pathogenic and symbiotic fungi," *Advances in Botanical Research*, vol. 49, pp. 151–193, 2009.

[25] G. Aguileta, G. Refrégier, R. Yockteng, E. Fournier, and T. Giraud, "Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists," *Infection, Genetics and Evolution*, vol. 9, no. 4, pp. 656–670, 2009.

[26] M. D. C. Alves and E. A. Pozza, "Scanning electron microscopy applied to seed-borne fungi examination," *Microscopy Research and Technique*, vol. 72, no. 7, pp. 482–488, 2009.

[27] V. Farkaš, "Structure and biosynthesis of fungal cell walls: methodological approaches," *Folia Microbiologica*, vol. 48, no. 4, pp. 469–478, 2003.

[28] A. R. Hardham and H. J. Mitchell, "Use of molecular cytology to study the structure and biology of phytopathogenic and mycorrhizal fungi," *Fungal Genetics and Biology*, vol. 24, no. 1-2, pp. 252–284, 1998.

[29] R. J. Howard, "Cytology of fungal pathogens and plant-host interactions," *Current Opinion in Microbiology*, vol. 4, no. 4, pp. 365–373, 2001.

[30] S. Koh and S. Somerville, "Show and tell: cell biology of pathogen invasion," *Current Opinion in Plant Biology*, vol. 9, no. 4, pp. 406–413, 2006.

[31] P. Pérez and J. C. Ribas, "Cell wall analysis," *Methods*, vol. 33, no. 3, pp. 245–251, 2004.

[32] J. A. Poland, P. J. Balint-Kurti, R. J. Wisser, R. C. Pratt, and R. J. Nelson, "Shades of gray: the world of quantitative disease resistance," *Trends in Plant Science*, vol. 14, no. 1, pp. 21–29, 2009.

[33] J. Xu, "Fundamentals of fungal molecular population genetic analyses," *Current Issues in Molecular Biology*, vol. 8, no. 2, pp. 75–89, 2006.

[34] M. Choquer, E. Fournier, C. Kunz, et al., "*Botrytis cinerea* virulence factors: new insights into a necrotrophic and polyphageous pathogen," *FEMS Microbiology Letters*, vol. 277, no. 1, pp. 1–10, 2007.

[35] M. J. Egan and N. J. Talbot, "Genomes, free radicals and plant cell invasion: recent developments in plant pathogenic fungi," *Current Opinion in Plant Biology*, vol. 11, no. 4, pp. 367–372, 2008.

[36] L. A. Hadwiger, "Localization predictions for gene products involved in non-host resistance responses in a model plant/fungal pathogen interaction," *Plant Science*, vol. 177, no. 4, pp. 257–265, 2009.

[37] K.-C. Tan, S. V. Ipcho, R. D. Trengove, R. P. Oliver, and P. S. Solomon, "Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology," *Molecular Plant Pathology*, vol. 10, no. 5, pp. 703–715, 2009.

[38] S. Walter, P. Nicholson, and F. M. Doohan, "Action and reaction of host and pathogen during *Fusarium* head blight disease," *New Phytologist*, vol. 185, no. 1, pp. 54–66, 2010.

[39] V. Bhadauria, S. Banniza, Y. Wei, and Y.-L. Peng, "Reverse genetics for functional genomics of phytopathogenic fungi and oomycetes," *Comparative and Functional Genomics*, vol. 2009, Article ID 380719, 11 pages, 2009.

[40] M. Sulc, K. Peslova, M. Zabka, M. Hajduch, and V. Havlicek, "Biomarkers of Aspergillus spores: strain typing and protein identification," *International Journal of Mass Spectrometry*, vol. 280, no. 1–3, pp. 162–168, 2009.

[41] M. L. Medina, P. A. Haynes, L. Breci, and W. A. Francisco, "Analysis of secreted proteins from *Aspergillus flavus*," *Proteomics*, vol. 5, no. 12, pp. 3153–3161, 2005.

[42] M. L. Medina, U. A. Kiernan, and W. A. Francisco, "Proteomic analysis of rutin-induced secreted proteins from *Aspergillus flavus*," *Fungal Genetics and Biology*, vol. 41, no. 3, pp. 327–335, 2004.

[43] S. Noir, T. Colby, A. Harzen, J. Schmidt, and R. Panstruga, "A proteomic analysis of powdery mildew (*Blumeria graminis* f.sp. hordei) conidiospores," *Molecular Plant Pathology*, vol. 10, no. 2, pp. 223–236, 2009.

[44] D. Godfrey, Z. Zhang, G. Saalbach, and H. Thordal-Christensen, "A proteomics study of barley powdery mildew haustoria," *Proteomics*, vol. 9, no. 12, pp. 3222–3232, 2009.

[45] F. J. Fernández-Acero, I. Jorge, E. Calvo, et al., "Two-dimensional electrophoresis protein profile of the phytopathogenic fungus *Botrytis cinerea*," *Proteomics*, vol. 6, supplement 1, pp. S88–S96, 2006.

[46] F. J. Fernández-Acero, I. Jorge, E. Calvo, et al., "Proteomic analysis of phytopathogenic fungus *Botrytis cinerea* as a potential tool for identifying pathogenicity factors, therapeutic targets and for basic research," *Archives of Microbiology*, vol. 187, no. 3, pp. 207–215, 2007.

[47] F. J. Fernández-Acero, T. Colby, A. Harzen, J. M. Cantoral, and J. Schmidt, "Proteomic analysis of the phytopathogenic fungus *Botrytis cinerea* during cellulose degradation," *Proteomics*, vol. 9, no. 10, pp. 2892–2902, 2009.

[48] P. Shah, J. A. Atwood, R. Orlando, H. E. Mubarek, G. K. Podila, and M. R. Davis, "Comparative proteomic analysis of *Botrytis cinerea* secretome," *Journal of Proteome Research*, vol. 8, no. 3, pp. 1123–1130, 2009.

[49] P. Shah, G. Gutierrez-Sanchez, R. Orlando, and C. Bergmann, "A proteomic study of pectin-degrading enzymes secreted by *Botrytis cinerea* grown in liquid culture," *Proteomics*, vol. 9, no. 11, pp. 3126–3135, 2009.

[50] S. Xu, J. I. E. Chen, L. Liu, X. Wang, X. Huang, and Y. Zhai, "Proteomics associated with virulence differentiation of *Curvularia iunata* in maize in China," *Journal of Integrative Plant Biology*, vol. 49, no. 4, pp. 487–496, 2007.

[51] V. Phalip, F. Delalande, C. Carapito, et al., "Diversity of the exoproteome of *Fusarium graminearum* grown on plant cell wall," *Current Genetics*, vol. 48, no. 6, pp. 366–379, 2005.

[52] J. M. Paper, J. S. Scott-Craig, N. D. Adhikari, C. A. Cuomo, and J. D. Walton, "Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus *Fusarium graminearum*," *Proteomics*, vol. 7, no. 17, pp. 3171–3183, 2007.

[53] R. D. Taylor, A. Saparno, B. Blackwell, et al., "Proteomic analyses of *Fusarium graminearum* grown under mycotoxin-inducing conditions," *Proteomics*, vol. 8, no. 11, pp. 2256–2265, 2008.

[54] S.-J. Kwon, S.-Y. Cho, K.-M. Lee, J. Yu, M. Son, and K.-H. Kim, "Proteomic analysis of fungal host factors differentially expressed by *Fusarium graminearum* infected with *Fusarium graminearum* virus-DK21," *Virus Research*, vol. 144, no. 1-2, pp. 96–106, 2009.

[55] X.-M. Zhao, X.-W. Zhang, W.-H. Tang, and L. Chen, "FPPI: *Fusarium graminearum* protein-protein interaction database," *Journal of Proteome Research*, vol. 8, no. 10, pp. 4714–4721, 2009.

[56] D. Vincent, M.-H. Balesdent, J. Gibon, et al., "Hunting down fungal secretomes using liquid-phase IEF prior to high resolution 2-DE," *Electrophoresis*, vol. 30, no. 23, pp. 4118–4136, 2009.

[57] F. He, Y. Zhang, H. Chen, Z. Zhang, and Y. L. Peng, "The prediction of protein-protein interaction networks in rice blast fungus," *BMC Genomics*, vol. 9, article 519, 2008.

[58] S. J. Shepherd, P. van West, and N. A. R. Gow, "Proteomic analysis of asexual development of *Phytophthora palmivora*," *Mycological Research*, vol. 107, no. 4, pp. 395–400, 2003.

[59] T. Ebstrup, G. Saalbach, and H. Egsgaard, "A proteomics study of in vitro cyst germination and appressoria formation in Phytophthora infestans," *Proteomics*, vol. 5, no. 11, pp. 2839–2848, 2005.

[60] H. J. G. Meijer, P. J. I. van de Vondervoort, Q. Y. Yin, et al., "Identification of cell wall-associated proteins from *Phytophthora ramorum*," *Molecular Plant-Microbe Interactions*, vol. 19, no. 12, pp. 1348–1358, 2006.

[61] T. Cao, Y. M. Kim, N. N. V. Kav, and S. E. Strelkov, "A proteomic evaluation of *Pyrenophora tritici-repentis*, causal agent of tan spot of wheat, reveals major differences between virulent and avirulent isolates," *Proteomics*, vol. 9, no. 5, pp. 1177–1196, 2009.

[62] M. Matis, M. Žakelj-Mavrič, and J. Peter-Katalinić, "Mass spectrometry and database search in the analysis of proteins from the fungus *Pleurotus ostreatus*," *Proteomics*, vol. 5, no. 1, pp. 67–75, 2005.

[63] D. K. Lakshman, S. S. Natarajan, S. Lakshman, W. M. Garrett, and A. K. Dhar, "Optimized protein extraction methods for proteomic analysis of *Rhizoctonia solani*," *Mycologia*, vol. 100, no. 6, pp. 867–875, 2008.

[64] W. Yajima and N. N. V. Kav, "The proteome of the phytopathogenic fungus *Sclerotinia sclerotiorum*," *Proteomics*, vol. 6, no. 22, pp. 5995–6007, 2006.

[65] K. A. R. C. Tan, J. L. Heazlewood, A. H. Millar, G. Thomson, R. P. Oliver, and P. S. Solomon, "A signaling-regulated, short-chain dehydrogenase of *Stagonospora nodorum* regulates asexual development," *Eukaryotic Cell*, vol. 7, no. 11, pp. 1916–1929, 2008.

[66] S. Bringans, J. K. Hane, T. Casey, et al., "Deep proteogenomics; high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins from the fungal wheat pathogen *Stagonospora nodorum*," *BMC Bioinformatics*, vol. 10, article 301, 2009.

[67] B. Cooper, A. Neelam, K. B. Campbell, et al., "Protein accumulation in the germinating *Uromyces appendiculatus* uredospore," *Molecular Plant-Microbe Interactions*, vol. 20, no. 7, pp. 857–866, 2007.

[68] M. Böhmer, T. Colby, C. Böhmer, A. Bräutigam, J. Schmidt, and M. Bölker, "Proteomic analysis of dimorphic transition in the phytopathogenic fungus Ustilago maydis," *Proteomics*, vol. 7, no. 5, pp. 675–685, 2007.

[69] R. J. Weld, K. M. Plummer, M. A. Carpenter, and H. J. Ridgway, "Approaches to functional genomics in filamentous fungi," *Cell Research*, vol. 16, no. 1, pp. 31–44, 2006.

[70] J. Schumacher, I. F. de Larrinoa, and B. Tudzynski, "Calcineurin-responsive zinc finger transcription factor CRZ1 of *Botrytis cinerea* is required for growth, development, and full virulence on bean plants," *Eukaryotic Cell*, vol. 7, no. 4, pp. 584–601, 2008.

[71] N. Ajiro, Y. Miyamoto, A. Masunaka, et al., "Role of the host-selective ACT-toxin synthesis gene ACTTS2 encoding an enoyl-reductase in pathogenicity of the tangerine pathotype of alternaria alternata," *Phytopathology*, vol. 100, no. 2, pp. 120–126, 2010.

[72] M.-H. Chi, S.-Y. Park, S. Kim, and Y.-H. Lee, "A novel pathogenicity gene is required in the rice blast fungus to suppress the basal defenses of the host," *PLoS Pathogens*, vol. 5, no. 4, Article ID e1000401, 2009.

[73] K. H. Lamour, L. Finley, O. Hurtado-Gonzales, D. Gobena, M. Tierney, and H. J. Meijer, "Targeted gene mutation in *Phytophthora* spp.," *Molecular Plant-Microbe Interactions*, vol. 19, no. 12, pp. 1359–1367, 2006.

[74] A. Di Pietro, M. P. Madrid, Z. Caracuel, J. Delgado-Jarana, and M. I. G. Roncero, "Fusarium oxysporum: exploring the molecular arsenal of a vascular wilt fungus," *Molecular Plant Pathology*, vol. 4, no. 5, pp. 315–325, 2003.

[75] S. R. Herron, J. A. Benen, R. D. Scavetta, J. Visser, and F. Jurnak, "Structure and function of pectic enzymes: virulence factors of plant pathogens," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 16, pp. 8762–8769, 2000.

[76] S. Meng, T. Torto-Alalibo, M. C. Chibucos, B. M. Tyler, and R. A. Dean, "Common processes in pathogenesis by fungal and oomycete plant pathogens, described with Gene Ontology terms," *BMC Microbiology*, vol. 9, supplement 1, p. S7, 2009.

[77] J. Delgado-Jarana, A. L. Martínez-Rocha, R. Roldán-Rodriguez, M. I. Roncero, and A. D. Pietro, "Fusarium oxysporum G-protein β subunit Fgb1 regulates hyphal growth, development, and virulence through multiple signalling pathways," *Fungal Genetics and Biology*, vol. 42, no. 1, pp. 61–72, 2005.

[78] N. Lee, C. A. D'Souza, and J. W. Kronstad, "Of smuts, blasts, mildews, and blights: cAMP signaling in phytopathogenic fungi," *Annual Review of Phytopathology*, vol. 41, pp. 399–427, 2003.

[79] N. Rispail and A. Di Pietro, "Fusarium oxysporum Ste12 controls invasive growth and virulence downstream of the Fmk1 MAPK cascade," *Molecular Plant-Microbe Interactions*, vol. 22, no. 7, pp. 830–839, 2009.

[80] J. R. Xu and J. E. Hamer, "MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*," *Genes and Development*, vol. 10, no. 21, pp. 2696–2706, 1996.

[81] L. Zheng, M. Campbell, J. Murphy, S. Lam, and J. R. Xu, "The BMP1 gene is essential for pathogenicity in the gray mold fungus *Botrytis cinerea*," *Molecular Plant-Microbe Interactions*, vol. 13, no. 7, pp. 724–732, 2000.

[82] N. Rispail, D. M. Soanes, C. Ant, et al., "Comparative genomics of MAP kinase and calcium-calcineurin signalling components in plant and human pathogenic fungi," *Fungal Genetics and Biology*, vol. 46, no. 4, pp. 287–298, 2009.

[83] M. S. López-Berges, A. Di Pietro, M. J. Daboussi, et al., "Identification of virulence genes in Fusarium oxysporum f. sp. lycopersici by large-scale transposon tagging," *Molecular Plant Pathology*, vol. 10, no. 1, pp. 95–107, 2009.

[84] V. Bhadauria, L. Popescu, W. S. Zhao, and Y. L. Peng, "Fungal transcriptomics," *Microbiological Research*, vol. 162, no. 4, pp. 285–298, 2007.

[85] Y. Oh, N. Donofrio, H. Pan, et al., "Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus Magnaporthe oryzae," *Genome Biology*, vol. 9, no. 5, R85, 2008.

[86] H. Takahara, A. Dolf, E. Endl, and R. O'Connell, "Flow cytometric purification of Colletotrichum higginsianum biotrophic hyphae from *Arabidopsis* leaves for stage-specific transcriptome analysis," *Plant Journal*, vol. 59, no. 4, pp. 672–683, 2009.

[87] R. P. Wise, M. J. Moscou, A. J. Bogdanove, and S. A. Whitham, "Transcript profiling in host-pathogen interactions," *Annual Review of Phytopathology*, vol. 45, pp. 329–369, 2008.

[88] B. Venkatesh, U. Hettwer, B. Koopmann, and P. Karlovsky, "Conversion of cDNA differential display results (DDRT-PCR) into quantitative transcription profiles," *BMC Genomics*, vol. 6, article 51, 2005.

[89] X. Wang, C. Tang, G. Zhang, et al., "cDNA-AFLP analysis reveals differential gene expression in compatible interaction of wheat challenged with Puccinia striiformis f. sp. tritici," *BMC Genomics*, vol. 10, article 289, 2009.

[90] C. Fekete, R. W. Fung, Z. Szabó, et al., "Up-regulated transcripts in a compatible powdery mildew-grapevine interaction," *Plant Physiology and Biochemistry*, vol. 47, no. 8, pp. 732–738, 2009.

[91] R. C. Venu, Y. Jia, M. Gowda, et al., "RL-SAGE and microarray analysis of the rice transcriptome after Rhizoctonia solani infection," *Molecular Genetics and Genomics*, vol. 278, no. 4, pp. 421–431, 2007.

[92] C. Yin, X. Chen, X. Wang, Q. Han, Z. Kang, and S. Hulbert, "Generation and analysis of expression sequence tags from haustoria of the wheat stripe rust fungus Puccinia striiformis f. sp. Tritici," *BMC Genomics*, vol. 10, article 626, 2009.

[93] W. Skinner, J. Keon, and J. Hargreaves, "Gene information for fungal plant pathogens from expressed sequences," *Current Opinion in Microbiology*, vol. 4, no. 4, pp. 381–386, 2001.

[94] F. J. Bruggeman and H. V. Westerhoff, "The nature of systems biology," *Trends in Microbiology*, vol. 15, no. 1, pp. 45–50, 2007.

[95] P. Picotti, B. Bodenmiller, L. N. Mueller, B. Domon, and R. Aebersold, "Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics," *Cell*, vol. 138, no. 4, pp. 795–806, 2009.

[96] N. L. Anderson, N. G. Anderson, T. W. Pearson, et al., "A human proteome detection and quantitation project," *Molecular and Cellular Proteomics*, vol. 8, no. 5, pp. 883–886, 2009.

[97] M. R. Wilkins, J. C. Sanchez, A. A. Gooley, et al., "Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it," *Biotechnology and Genetic Engineering Reviews*, vol. 13, pp. 19–50, 1996.

[98] J. Cox and M. Mann, "Is proteomics the new genomics?" *Cell*, vol. 130, no. 3, pp. 395–398, 2007.

[99] B. F. Cravatt, G. M. Simon, and J. R. Yates III, "The biological impact of mass-spectrometry-based proteomics," *Nature*, vol. 450, no. 7172, pp. 991–1000, 2007.

[100] X. Han, M. Jin, K. Breuker, and F. W. McLafferty, "Extending top-down mass spectrometry to proteins with masses great than 200 kilodaltons," *Science*, vol. 314, no. 5796, pp. 109–112, 2006.

[101] X. Han, A. Aslanian, and J. R. Yates III, "Mass spectrometry for proteomics," *Current Opinion in Chemical Biology*, vol. 12, no. 5, pp. 483–490, 2008.

[102] J. V. Jorrín-Novo, A. M. Maldonado, S. Echevarría-Zomeño, et al., "Plant proteomics update (2007-2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge," *Journal of Proteomics*, vol. 72, no. 3, pp. 285–314, 2009.

[103] M. Mann, "Comparative analysis to guide quality improvements in proteomics," *Nature Methods*, vol. 6, no. 10, pp. 717–719, 2009.

[104] A. Schmidt, M. Claassen, and R. Aebersold, "Directed mass spectrometry: towards hypothesis-driven proteomics," *Current Opinion in Chemical Biology*, vol. 13, no. 5-6, pp. 510–517, 2009.

[105] V. Bhadauria, W. S. Zhao, L. X. Wang, et al., "Advances in fungal proteomics," *Microbiological Research*, vol. 162, no. 3, pp. 193–200, 2007.

[106] Y. Kim, M. P. Nandakumar, and M. R. Marten, "Proteomics of filamentous fungi," *Trends in Biotechnology*, vol. 25, no. 9, pp. 395–400, 2007.

[107] L. J. Newey, C. E. Caten, and J. R. Green, "Rapid adhesion of Stagonospora nodorum spores to a hydrophobic surface requires pre-formed cell surface glycoproteins," *Mycological Research*, vol. 111, no. 11, pp. 1255–1267, 2007.

[108] P. N. Dodds, M. Rafiqi, P. H. P. Gan, A. R. Hardham, D. A. Jones, and J. G. Ellis, "Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance," *New Phytologist*, vol. 183, no. 4, pp. 993–1000, 2009.

[109] P. M. Houterman, B. J. Cornelissen, and M. Rep, "Suppression of plant resistance gene-based immunity by a fungal effector," *PLoS Pathogens*, vol. 4, no. 5, Article ID e1000061, 2008.

[110] J. Song, J. Win, M. Tian, et al., "Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 5, pp. 1654–1659, 2009.

[111] P. J. G. M. De Wit, M. B. Buurlage, and K. E. Hammond, "The occurrence of host-, pathogen- and interaction-specific proteins in the apoplast of *Cladosporium fulvum* (syn. *Fulvia fulva*) infected tomato leaves," *Physiological and Molecular Plant Pathology*, vol. 29, no. 2, pp. 159–172, 1986.

[112] I. M. J. Schottens-Toma and P. J. G. M. De Wit, "Purification and primary structure of a necrosis-inducing peptide from the apoplastic fluids of tomato infected with *Cladosporium fulvum* (syn. *Fulvia fulva*)," *Physiological and Molecular Plant Pathology*, vol. 33, no. 1, pp. 59–67, 1988.

[113] J. G. Ellis, M. Rafiqi, P. Gan, A. Chakrabarti, and P. N. Dodds, "Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens," *Current Opinion in Plant Biology*, vol. 12, no. 4, pp. 399–405, 2009.

[114] M. Rep, H. C. van der Does, M. Meijer, et al., "A small, cysteine-rich protein secreted by Fusarium oxysporum during colonization of xylem vessels is required for I-3-mediated resistance in tomato," *Molecular Microbiology*, vol. 53, no. 5, pp. 1373–1383, 2004.

[115] M. Rep, "Small proteins of plant-pathogenic fungi secreted during host colonization," *FEMS Microbiology Letters*, vol. 253, no. 1, pp. 19–27, 2005.

[116] E. W. Deutsch, H. Lam, and R. Aebersold, "Data analysis and bioinformatics tools for tandem mass spectrometry in

proteomics," *Physiological Genomics*, vol. 33, no. 1, pp. 18–25, 2008.

[117] J. Eriksson and D. Fenyö, "Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs," *Nature Biotechnology*, vol. 25, no. 6, pp. 651–655, 2007.

[118] P. Holzmuller, P. Grébaut, J. P. Brizard, et al., "Pathogeno-proteomics," *Annals of the New York Academy of Sciences*, vol. 1149, pp. 66–70, 2008.

[119] L. Valledor, M. A. Castillejo, C. Lenz, R. Rodríguez, M. J. Cañal, and J. Jorrín, "Proteomic analysis of Pinus radiata needles: 2-DE map and protein identification by LC/MS/MS and substitution-tolerant database searching," *Journal of Proteome Research*, vol. 7, no. 7, pp. 2616–2631, 2008.

[120] J. Ruiz-Herrera, *Fungal Cell Wall: Structure, Synthesis and Assembly*, CRC Press, Boca Raton, Fla, USA, 1992.

[121] J. Grinyer, M. McKay, B. Herbert, and H. Nevalainen, "Fungal proteomics: mapping the mitochodrial proteins of a Trichoderma harzianum strain applied for biological control," *Current Genetics*, vol. 45, no. 3, pp. 170–175, 2004.

[122] P. Melin, J. Schnürer, and E. G. Wagner, "Proteome analysis of Aspergillus nidulans reveals proteins associated with the response to the antibiotic concanamycin A, produced by Streptomyces species," *Molecular Genetics and Genomics*, vol. 267, no. 6, pp. 695–702, 2002.

[123] M. P. Nandakumar, J. Shen, B. Raman, and M. R. Marten, "Solubilization of trichloroacetic acid (TCA) precipitated microbial proteins via NaOH for two-dimensional electrophoresis," *Journal of Proteome Research*, vol. 2, no. 1, pp. 89–93, 2003.

[124] K. Ström, J. Schnürer, and P. Melin, "Co-cultivation of anti-fungal Lactobacillus plantarum MiLAB 393 and Aspergillus nidulans, evaluation of effects on fungal growth and protein expression," *FEMS Microbiology Letters*, vol. 246, no. 1, pp. 119–124, 2005.

[125] J. M. Bruneau, T. Magnin, E. Tagat, et al., "Proteome analysis of Aspergillus fumigatus identifies glycosylphosphat-idylinositol-anchored proteins associated to the cell wall biosynthesis," *Electrophoresis*, vol. 22, no. 13, pp. 2812–2823, 2001.

[126] J. Grinyer, S. Hunt, M. McKay, B. E. N. R. Herbert, and H. Nevalainen, "Proteomic response of the biological control fungus Trichoderma atroviride to growth on the cell walls of *Rhizoctonia solani*," *Current Genetics*, vol. 47, no. 6, pp. 381–388, 2005.

[127] J. Grinyer, M. McKay, H. Nevalainen, and B. E. N. R. Herbert, "Fungal proteomics: initial mapping of biological control strain Trichoderma harzianum," *Current Genetics*, vol. 45, no. 3, pp. 163–169, 2004.

[128] M. P. Nandakumar and M. R. Marten, "Comparison of lysis methods and preparation protocols for one- and two-dimensional electrophoresis of *Aspergillus oryzae* intracellu-lar proteins," *Electrophoresis*, vol. 23, no. 14, pp. 2216–2222, 2002.

[129] M. Shimizu and H. Wariishi, "Development of a sample preparation method for fungal proteomics," *FEMS Microbi-ology Letters*, vol. 247, no. 1, pp. 17–22, 2005.

[130] M. Shimizu, N. Yuda, T. Nakamura, H. Tanaka, and H. Wariishi, "Metabolic regulation at the tricarboxylic acid and glyoxylate cycles of the lignin-degrading basidiomycete *Phanerochaete chrysosporium* against exogenous addition of vanillin," *Proteomics*, vol. 5, no. 15, pp. 3919–3931, 2005.

[131] M. L. Hernández-Macedo, A. Ferraz, J. Rodrguez, L. M. M. Ottoboni, and M. P. De Mello, "Iron-regulated proteins in *Phanerochaete chrysosporium* and *Lentinula edodes*: dif-ferential analysis by sodium dodecyl sulfate polyacrylamide gel electrophoresis and two-dimensional polyacrylamide gel electrophoresis profiles," *Electrophoresis*, vol. 23, no. 4, pp. 655–661, 2002.

[132] A. Görg, W. Weiss, and M. J. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, 2004.

[133] S. C. Carpentier, E. Witters, K. Laukens, P. Deckers, R. Swennen, and B. Panis, "Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis," *Proteomics*, vol. 5, no. 10, pp. 2497–2507, 2005.

[134] C. Damerval, D. de Vienne, M. Zivy, and H. Thiellement, "Technical improvements in two-dimensional electrophore-sis increase the level of genetic variation detected in wheat-seedling proteins," *Electrophoresis*, vol. 7, pp. 52–54, 1986.

[135] A. N. A. M. Maldonado, S. Echevarría-Zomeño, S. Jean-Baptiste, M. Hernández, and J. V. Jorrín-Novo, "Evaluation of three different protocols of protein extraction for *Ara-bidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis," *Journal of Proteomics*, vol. 71, no. 4, pp. 461–472, 2008.

[136] R. Wildgruber, G. Reil, O. Drews, H. Parlar, and A. Görg, "Web-based two-dimensional database of *Saccharomyces cerevisiae* proteins using immobilized pH gradients from pH 6 to pH 12 and matrix-assisted laser desorption/ionization-time of flight mass spectrometry," *Proteomics*, vol. 2, no. 6, pp. 727–732, 2002.

[137] H. Everberg, N. Gustavsson, and F. Tjerned, "Enrichment of membrane proteins by partitioning in detergent/polymer aqueous two-phase systems," *Methods in Molecular Biology*, vol. 424, pp. 403–412, 2008.

[138] S. Luche, V. Santoni, and T. Rabilloud, "Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis," *Proteomics*, vol. 3, no. 3, pp. 249–253, 2003.

[139] M. P. Molloy, B. R. Herbert, K. L. Williams, and A. A. Gooley, "Extraction of *Escherichia coli* proteins with organic solvents prior to two-dimensional electrophoresis," *Electrophoresis*, vol. 20, no. 4-5, pp. 701–704, 1999.

[140] T. Rabilloud, "Solubilization of proteins for electrophoretic analyses," *Electrophoresis*, vol. 17, no. 5, pp. 813–829, 1996.

[141] T. Rabilloud, C. Adessi, A. Giraudel, and J. Lunardi, "Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradi-ents," *Electrophoresis*, vol. 18, no. 3-4, pp. 307–316, 1997.

[142] O. Kniemeyer, F. Lessing, O. Scheibner, C. Hertweck, and A. A. Brakhage, "Optimisation of a 2-D gel electrophoresis protocol for the human-pathogenic fungus *Aspergillus fumi-gatus*," *Current Genetics*, vol. 49, no. 3, pp. 178–189, 2006.

[143] T. Rabilloud, "Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis," *Electrophoresis*, vol. 19, no. 5, pp. 758–760, 1998.

[144] B. E. N. R. Herbert, J. Grinyer, J. T. McCarthy, et al., "Improved 2-DE of microorganisms after acidic extraction," *Electrophoresis*, vol. 27, no. 8, pp. 1630–1640, 2006.

[145] O. Guais, G. Borderies, C. Pichereaux, et al., "Proteomics analysis of "Rovabiot Excel", a secreted protein cocktail from the filamentous fungus Penicillium funiculosum grown under industrial process fermentation," *Journal of Industrial Microbiology & Biotechnology* , pp. 1659–1668, 2008.

[146] S. H. Kao, H. I. N. K. Wong, C. Y. Chiang, and H. A. N. M. Chen, "Evaluating the compatibility of three colorimetric

protein assays for two-dimensional electrophoresis experiments," *Proteomics*, vol. 8, no. 11, pp. 2178–2184, 2008.

[147] D. Fragner, M. Zomorrodi, U. Kües, and A. Majcherczyk, "Optimized protocol for the 2-DE of extracellular proteins from higher basidiomycetes inhabiting lignocellulose," *Electrophoresis*, vol. 30, no. 14, pp. 2431–2441, 2009.

[148] M. L. Medina and W. A. Francisco, "Isolation and enrichment of secreted proteins from filamentous fungi," in *2D PAGE: Sample Preparation and Fractionation*, pp. 275–285, Springer, New York, NY, USA, 2008.

[149] A. Abbas, H. Koc, F. Liu, and M. Tien, "Fungal degradation of wood: initial proteomic analysis of extracellular proteins of *Phanerochaete chrysosporium* grown on oak substrate," *Current Genetics*, vol. 47, no. 1, pp. 49–56, 2005.

[150] H. Ravalason, G. Jan, D. Mollé, et al., "Secretome analysis of *Phanerochaete chrysosporium* strain CIRM-BRFM41 grown on softwood," *Applied Microbiology and Biotechnology*, vol. 80, no. 4, pp. 719–733, 2008.

[151] A. V. Wymelenberg, P. Minges, G. Sabat, et al., "Computational analysis of the *Phanerochaete chrysosporium* v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins," *Fungal Genetics and Biology*, vol. 43, no. 5, pp. 343–356, 2006.

[152] H. Zorn, T. Peters, M. Nimtz, and R. G. Berger, "The secretome of *Pleurotus sapidus*," *Proteomics*, vol. 5, no. 18, pp. 4832–4838, 2005.

[153] L. V. Bindschedler, T. A. Burgis, D. J. S. Mills, J. T. C. Ho, R. Cramer, and P. D. Spanu, "In planta proteomics and proteogenomics of the biotrophic Barley fungal pathogen *Blumeria graminis* f. sp. *hordei*," *Molecular and Cellular Proteomics*, vol. 8, no. 10, pp. 2368–2381, 2009.

[154] C. Rampitsch, N. V. Bykova, B. McCallum, E. V. A. Beimcik, and W. Ens, "Analysis of the wheat and Puccinia triticina (leaf rust) proteomes during a susceptible host-pathogen interaction," *Proteomics*, vol. 6, no. 6, pp. 1897–1907, 2006.

[155] W. Zhou, F. Eudes, and A. Laroche, "Identification of differentially regulated proteins in response to a compatible interaction between the pathogen *Fusarium graminearum* and its host, *Triticum aestivum*," *Proteomics*, vol. 6, no. 16, pp. 4599–4609, 2006.

[156] P. M. Houterman, D. Speijer, H. L. Dekker, C. G. de Koster, B. J. C. Cornelissen, and M. Rep, "The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants," *Molecular Plant Pathology*, vol. 8, no. 2, pp. 215–221, 2007.

[157] A. Harder, "Sample preparation procedure for cellular fungi," *Methods in Molecular Biology*, vol. 425, pp. 265–273, 2008.

[158] A. Pitarch, C. Nombela, and C. Gil, "Cell wall fractionation for yeast and fungal proteomics," *Methods in Molecular Biology*, vol. 425, pp. 217–239, 2008.

[159] X. Jiang, M. Ye, and H. Zou, "Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis," *Proteomics*, vol. 8, no. 4, pp. 686–705, 2008.

[160] E. Boschetti and P. G. Righetti, "The art of observing rare protein species in proteomes with peptide ligand libraries," *Proteomics*, vol. 9, no. 6, pp. 1492–1510, 2009.

[161] D. Robertson, G. P. Mitchell, J. S. Gilroy, C. Gerrish, G. P. Bolwell, and A. R. Slabas, "Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants," *Journal of Biological Chemistry*, vol. 272, no. 25, pp. 15841–15848, 1997.

[162] F. Supek, P. Peharec, M. Krsnik-Rasol, and T. Šmuc, "Enhanced analytical power of SDS-PAGE using machine learning algorithms," *Proteomics*, vol. 8, no. 1, pp. 28–31, 2008.

[163] L. M. F. de Godoy, J. V. Olsen, G. A. de Souza, G. Li, P. Mortensen, and M. Mann, "Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system," *Genome Biology*, vol. 7, no. 6, article R50, 2006.

[164] F. Tribl, C. Lohaus, T. Dombert, et al., "Towards multidimensional liquid chromatography separation of proteins using fluorescence and isotope-coded protein labelling for quantitative proteomics," *Proteomics*, vol. 8, no. 6, pp. 1204–1211, 2008.

[165] B. G. Fryksdale, P. T. Jedrzejewski, D. L. Wong, A. L. Gaertner, and B. S. Miller, "Impact of deglycosylation methods on two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry for proteomic analysis," *Electrophoresis*, vol. 23, no. 14, pp. 2184–2193, 2002.

[166] A. V. Wymelenberg, G. Sabat, M. Mozuch, P. J. Kersten, D. A. N. Cullen, and R. A. Blanchette, "Structure, organization, and transcriptional regulation of a family of copper radical oxidase genes in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*," *Applied and Environmental Microbiology*, vol. 72, no. 7, pp. 4871–4877, 2006.

[167] P. Jungblut and B. Thiede, "Protein identification from 2-DE gels by MALDI mass spectrometry," *Mass Spectrometry Reviews*, vol. 16, no. 3, pp. 145–162, 1997.

[168] P. H. O'Farrell, "High resolution two dimensional electrophoresis of proteins," *Journal of Biological Chemistry*, vol. 250, no. 10, pp. 4007–4021, 1975.

[169] J. Klose and U. Kobalz, "Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome," *Electrophoresis*, vol. 16, no. 6, pp. 1034–1059, 1995.

[170] T. Rabilloud, A. R. Vaezzadeh, N. Potier, C. Lelong, E. Leize-Wagner, and M. Chevallet, "Power and limitations of electrophoretic separations in proteomics strategies," *Mass Spectrometry Reviews*, vol. 28, no. 5, pp. 816–843, 2009.

[171] P. A. Haynes and T. H. Roberts, "Subcellular shotgun proteomics in plants: looking beyond the usual suspects," *Proteomics*, vol. 7, no. 16, pp. 2963–2975, 2007.

[172] A. Pirondini, G. Visioli, A. Malcevschi, and N. Marmiroli, "A 2-D liquid-phase chromatography for proteomic analysis in plant tissues," *Journal of Chromatography B*, vol. 833, no. 1, pp. 91–100, 2006.

[173] G. Recorbet, H. Rogniaux, V. Gianinazzi-Pearson, and E. DumasGaudot, "Fungal proteins in the extra-radical phase of arbuscular mycorrhiza: a shotgun proteomic picture," *New Phytologist*, vol. 181, no. 2, pp. 248–260, 2009.

[174] M. Ye, X. Jiang, S. Feng, R. Tian, and H. Zou, "Advances in chromatographic techniques and methods in shotgun proteome analysis," *Trends in Analytical Chemistry*, vol. 26, no. 1, pp. 80–84, 2007.

[175] M. P. Washburn, D. Wolters, and J. R. Yates III, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology," *Nature Biotechnology*, vol. 19, no. 3, pp. 242–247, 2001.

[176] T. McDonald, S. Sheng, B. Stanley, et al., "Expanding the subproteome of the inner mitochondria using protein separation technologies: one-and two-dimensional liquid chromatography and two-dimensional gel electrophoresis," *Molecular and Cellular Proteomics*, vol. 5, no. 12, pp. 2392–2411, 2006.

[177] W. W. Wu, G. Wang, S. J. Baek, and R. F. Shen, "Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF," *Journal of Proteome Research*, vol. 5, no. 3, pp. 651–658, 2006.

[178] B. Domon and R. Aebersold, "Mass spectrometry and protein analysis," *Science*, vol. 312, no. 5771, pp. 212–217, 2006.

[179] A. I. Nesvizhskii, O. Vitek, and R. Aebersold, "Analysis and validation of proteomic data generated by tandem mass spectrometry," *Nature Methods*, vol. 4, no. 10, pp. 787–797, 2007.

[180] D. M. Soanes, W. Skinner, J. Keon, J. Hargreaves, and N. J. Talbot, "Genomics of phytopathogenic fungi and the development of bioinformatic resources," *Molecular Plant-Microbe Interactions*, vol. 15, no. 5, pp. 421–427, 2002.

[181] D. M. Soanes and N. J. Talbot, "Comparative genomic analysis of phytopathogenic fungi using expressed sequence tag (EST) collections," *Molecular Plant Pathology*, vol. 7, no. 1, pp. 61–70, 2006.

[182] J. Fiévet, C. Dillmann, G. Lagniel, et al., "Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis," *Proteomics*, vol. 4, no. 7, pp. 1939–1949, 2004.

[183] G. B. Smejkal, M. H. Robinson, and A. Lazarev, "Comparison of fluorescent stains: relative photostability and differential staining of proteins in two-dimensional gels," *Electrophoresis*, vol. 25, no. 15, pp. 2511–2519, 2004.

[184] Y. I. Hu, G. Wang, G. Y. J. Chen, X. I. N. Fu, and S. Q. Yao, "Proteome analysis of *Saccharomyces cerevisiae* under metal stress by two-dimensional differential gel electrophoresis," *Electrophoresis*, vol. 24, no. 9, pp. 1458–1470, 2003.

[185] J. U. N. X. Yan, A. T. Devenish, R. Wait, T. I. M. Stone, S. Lewis, and S. U. E. Fowler, "Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of *Escherichia coli*," *Proteomics*, vol. 2, no. 12, pp. 1682–1698, 2002.

[186] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags," *Nature Biotechnology*, vol. 17, no. 10, pp. 994–999, 1999.

[187] H. Zhou, J. A. Ranish, J. D. Watts, and R. Aebersold, "Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry," *Nature Biotechnology*, vol. 20, no. 5, pp. 512–515, 2002.

[188] X. Yao, A. Freas, J. Ramirez, P. A. Demirev, and C. Fenselau, "Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus," *Analytical Chemistry*, vol. 73, no. 13, pp. 2836–2842, 2001.

[189] P. A. Everley, J. Krijgsveld, B. R. Zetter, and S. P. Gygi, "Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research," *Molecular and Cellular Proteomics*, vol. 3, no. 7, pp. 729–735, 2004.

[190] L. R. Zieske, "A perspective on the use of iTRAQ^{TM} reagent technology for protein complex and profiling studies," *Journal of Experimental Botany*, vol. 57, no. 7, pp. 1501–1508, 2006.

[191] W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, et al., "Comparison of label-free methods for quantifying human proteins by shotgun proteomics," *Molecular and Cellular Proteomics*, vol. 4, no. 10, pp. 1487–1502, 2005.

[192] B. Zhang, N. C. VerBerkmoes, M. A. Langston, E. Uberbacher, R. L. Hettich, and N. F. Samatova, "Detecting differential and correlated protein expression in label-free shotgun proteomics," *Journal of Proteome Research*, vol. 5, no. 11, pp. 2909–2918, 2006.

[193] M. Ünlü, M. E. Morgan, and J. S. Minden, "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts," *Electrophoresis*, vol. 18, no. 11, pp. 2071–2077, 1997.

[194] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative mass spectrometry in proteomics: a critical review," *Analytical and Bioanalytical Chemistry*, vol. 389, no. 4, pp. 1017–1031, 2007.

[195] L. N. Mueller, M. I. Y. Brusniak, D. R. Mani, and R. Aebersold, "An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data," *Journal of Proteome Research*, vol. 7, no. 1, pp. 51–61, 2008.

[196] V. J. Patel, K. Thalassinos, S. E. Slade, et al., "A comparison of labeling and label-free mass spectrometry-based proteomics approaches," *Journal of Proteome Research*, vol. 8, no. 7, pp. 3752–3759, 2009.

[197] S. Xie, C. Moya, B. Bilgin, A. Jayaraman, and S. P. Walton, "Emerging affinity-based techniques in proteomics," *Expert Review of Proteomics*, vol. 6, no. 5, pp. 573–583, 2009.

[198] I. A. N. P. Shadforth, T. P. J. Dunkley, K. S. Lilley, and C. Bessant, "i-Tracker: for quantitative proteomics using iTRAQ^{TM}," *BMC Genomics*, vol. 6, article 145, 2005.

[199] P. V. Bondarenko, D. Chelius, and T. A. Shaler, "Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry," *Analytical Chemistry*, vol. 74, no. 18, pp. 4741–4749, 2002.

[200] D. Chelius and P. V. Bondarenko, "Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry," *Journal of Proteome Research*, vol. 1, no. 4, pp. 317–323, 2002.

[201] W. Wang, H. Zhou, H. U. A. Lin, et al., "Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards," *Analytical Chemistry*, vol. 75, no. 18, pp. 4818–4826, 2003.

[202] H. Liu, R. G. Sadygov, and J. R. Yates III, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics," *Analytical Chemistry*, vol. 76, no. 14, pp. 4193–4201, 2004.

[203] J. Fasolo and M. Snyder, "Protein microarrays," *Methods in Molecular Biology*, vol. 548, pp. 209–222, 2009.

[204] J. Barrett, P. M. Brophy, and J. V. Hamilton, "Analysing proteomic data," *International Journal for Parasitology*, vol. 35, no. 5, pp. 543–553, 2005.

[205] E. Marengo, E. Robotti, F. Antonucci, D. Cecconi, N. Campostrini, and P. G. Righetti, "Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems," *Proteomics*, vol. 5, no. 3, pp. 654–666, 2005.

[206] Å. M. Wheelock and S. Goto, "Effects of post-electrophoretic analysis on variance in gel-based proteomics," *Expert Review of Proteomics*, vol. 3, no. 1, pp. 129–142, 2006.

[207] E. W. Deutsch, H. Lam, and R. Aebersold, "PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows," *EMBO Reports*, vol. 9, no. 5, pp. 429–434, 2008.

[208] D. G. Biron, C. Brun, T. Lefevre, et al., "The pitfalls of proteomics experiments without the correct use of bioinformatics tools," *Proteomics*, vol. 6, no. 20, pp. 5577–5596, 2006.

[209] N. A. Karp and K. S. Lilley, "Design and analysis issues in quantitative proteomics studies," *Proteomics*, vol. 2, no. 1, pp. 42–50, 2007.

[210] J. M. Bland and D. G. Altman, "Multiple significance tests: the Bonferroni method," *British Medical Journal*, vol. 310, no. 6973, article 170, 1995.

[211] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B*, vol. 57, pp. 289–300, 1995.

[212] J. D. Storey and R. Tibshirani, "Statistical significance for genomewide studies," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9440–9445, 2003.

[213] K. Strimmer, "fdrtool: a versatile R package for estimating local and tail area-based false discovery rates," *Bioinformatics*, vol. 24, no. 12, pp. 1461–1462, 2008.

[214] S. Jacobsen, H. Grove, K. N. Jensen, et al., "Multivariate analysis of 2-DE protein patterns—practical approaches," *Electrophoresis*, vol. 28, no. 8, pp. 1289–1299, 2007.

[215] K. N. Jensen, F. Jessen, and B. O. M. Jørgensen, "Multivariate data analysis of two-dimensional gel electrophoresis protein patterns from few samples," *Journal of Proteome Research*, vol. 7, no. 3, pp. 1288–1296, 2008.

[216] R. Pedreschi, M. L. A. T. M. Hertog, S. C. Carpentier, et al., "Treatment of missing values for multivariate statistical analysis of gel-based proteomics data," *Proteomics*, vol. 8, no. 7, pp. 1371–1383, 2008.

[217] M. Schneider, L. Lane, E. Boutet, et al., "The UniProtKB/ Swiss-Prot knowledgebase and its Plant Proteome Annotation Program," *Journal of Proteomics*, vol. 72, no. 3, pp. 567–573, 2009.

[218] C. F. Taylor, N. W. Paton, K. S. Lilley, et al., "The minimum information about a proteomics experiment (MIAPE)," *Nature Biotechnology*, vol. 25, no. 8, pp. 887–893, 2007.

[219] J. A. Mead, I. A. N. P. Shadforth, and C. Bessant, "Public proteomic MS repositories and pipelines: available tools and biological applications," *Proteomics*, vol. 7, no. 16, pp. 2769–2786, 2007.

[220] M. Fälth, M. M. Savitski, M. L. Nielsen, F. Kjeldsen, P. E. R. E. Andren, and R. A. Zubarev, "SwedCAD, a database of annotated high-mass accuracy MS/MS spectra of tryptic peptides," *Journal of Proteome Research*, vol. 6, no. 10, pp. 4063–4067, 2007.

[221] J. A. N. Hummel, M. Niemann, S. Wienkoop, et al., "ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites," *BMC Bioinformatics*, vol. 8, article 216, 2007.

[222] H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, S. E. Stein, and R. Aebersold, "Building consensus spectral libraries for peptide identification in proteomics," *Nature Methods*, vol. 5, no. 10, pp. 873–875, 2008.

[223] R. Craig, J. P. Cortens, and R. C. Beavis, "Open source system for analyzing, validating, and storing protein identification data," *Journal of Proteome Research*, vol. 3, no. 6, pp. 1234–1242, 2004.

[224] F. Desiere, E. W. Deutsch, N. L. King, et al., "The peptide atlas project," *Nucleic Acids Research*, vol. 34, pp. D655–658, 2006.

[225] L. Martens, H. Hermjakob, P. Jones, et al., "PRIDE: the proteomics identifications database," *Proteomics*, vol. 5, no. 13, pp. 3537–3545, 2005.

[226] I. A. N. Shadforth, W. Xu, D. Crowther, and C. Bessant, "GAPP: a fully automated software for the confident identification of human peptides from tandem mass spectra," *Journal of Proteome Research*, vol. 5, no. 10, pp. 2849–2852, 2006.

[227] T. McLaughlin, J. A. Siepen, J. Selley, et al., "PepSeeker: a database of proteome peptide identifications for investigating fragmentation patterns," *Nucleic Acids Research*, vol. 34, pp. D649–D654, 2006.

[228] Y. Zhang, Y. Zhang, J. U. N. Adachi, et al., "MAPU: Max-Planck Unified database of organellar, cellular, tissue and body fluid proteomes," *Nucleic Acids Research*, vol. 35, database issue, pp. D771–D779, 2007.

[229] J. T. Prince, M. W. Carlson, R. Wang, P. Lu, and E. M. Marcotte, "The need for a public proteomics repository," *Nature Biotechnology*, vol. 22, no. 4, pp. 471–472, 2004.

[230] M. Riffle, L. Malmström, and T. N. Davis, "The yeast resource center public data repository," *Nucleic Acids Research*, vol. 33, pp. D378–D382, 2005.

[231] I. L. S. Oh, A. E. R. A. N. Park, M. I. N. S. Bae, et al., "Secretome analysis reveals an *Arabidopsis* lipase involved in defense against Altemaria brassicicola," *Plant Cell*, vol. 17, no. 10, pp. 2832–2847, 2005.

[232] F. Colditz, O. Nyamsuren, K. Niehaus, H. Eubel, H. P. Braun, and F. Krajinski, "Proteomic approach: identification of *Medicago truncatula* proteins induced in roots after infection with the pathogenic oomycete *Aphanomyces euteiches*," *Plant Molecular Biology*, vol. 55, no. 1, pp. 109–120, 2004.

[233] F. Colditz, H. P. Braun, C. Jacquet, K. Niehaus, and F. Krajinski, "Proteomic profiling unravels insights into the molecular background underlying increased *Aphanomyces euteiches*-tolerance of *Medicago truncatula*," *Plant Molecular Biology*, vol. 59, no. 3, pp. 387–406, 2005.

[234] F. Colditz, K. Niehaus, and F. Krajinski, "Silencing of PR-10-like proteins in *Medicago truncatula* results in an antagonistic induction of other PR proteins and in an increased tolerance upon infection with the oomycete *Aphanomyces euteiches*," *Planta*, vol. 226, no. 1, pp. 57–71, 2007.

[235] T. J. March, J. A. Able, C. J. Schultz, and A. J. Able, "A novel late embryogenesis abundant protein and peroxidase associated with black point in barley grains," *Proteomics*, vol. 7, no. 20, pp. 3800–3808, 2007.

[236] M. D. Bolton, H. P. van Esse, J. H. Vossen, et al., "The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species," *Molecular Microbiology*, vol. 69, no. 1, pp. 119–136, 2008.

[237] J. A. Smith, R. A. Blanchette, T. A. Burnes, et al., "Proteomic comparison of needles from blister rust-resistant and susceptible Pinus strobus seedlings reveals upregulation of putative disease resistance proteins," *Molecular Plant-Microbe Interactions*, vol. 19, no. 2, pp. 150–160, 2006.

[238] D. Wang, A. Eyles, D. Mandich, and P. Bonello, "Systemic aspects of host-pathogen interactions in Austrian pine (Pinus nigra): a proteomics approach," *Physiological and Molecular Plant Pathology*, vol. 68, no. 4–6, pp. 149–157, 2006.

[239] M. Curto, E. Camafeita, J. A. Lopez, A. N. A. M. Maldonado, D. Rubiales, and J. V. Jorrín, "A proteomic approach to study pea (Pisum sativum) responses to powdery mildew (Erysiphe pisi)," *Proteomics*, vol. 6, supplement 1, pp. S163–S174, 2006.

[240] J. Geddes, F. Eudes, A. Laroche, and L. B. Selinger, "Differential expression of proteins in response to the interaction between the pathogen *Fusarium graminearum* and its host, Hordeum vulgare," *Proteomics*, vol. 8, no. 3, pp. 545–554, 2008.

[241] W. Zhou, F. L. Kolb, and D. E. Riechers, "Identification of proteins induced or upregulated by Fusarium head blight

infection in the spikes of hexaploid wheat (*Triticum aestivum*)," *Genome*, vol. 48, no. 5, pp. 770–780, 2005.

[242] B. K. Ndimba, S. Chivasa, J. M. Hamilton, W. J. Simon, and A. R. Slabas, "Proteomic analysis of changes in the extracellular matrix of *Arabidopsis* cell suspension cultures induced by fungal elicitors," *Proteomics*, vol. 3, no. 6, pp. 1047–1059, 2003.

[243] R. L. Larson, A. M. Y. L. Hill, and A. Nuñez, "Characterization of protein changes associated with sugar beet (*Beta vulgaris*) resistance and susceptibility to Fusarium oxysporum," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 19, pp. 7905–7915, 2007.

[244] S. Campo, M. Carrascal, M. Coca, J. Abián, and B. San Segundo, "The defense response of germinating maize embryos against fungal infection: a proteomics approach," *Proteomics*, vol. 4, no. 2, pp. 383–396, 2004.

[245] S. Chivasa, J. M. Hamilton, R. S. Pringle, et al., "Proteomic analysis of differentially expressed proteins in fungal elicitor-treated *Arabidopsis* cell cultures," *Journal of Experimental Botany*, vol. 57, no. 7, pp. 1553–1562, 2006.

[246] S. Chivasa, W. J. Simon, X. L. Yu, N. Yalpani, and A. R. Slabas, "Pathogen elicitor-induced changes in the maize extracellular matrix proteome," *Proteomics*, vol. 5, no. 18, pp. 4894–4904, 2005.

[247] J. V. F. Coumans, A. Poljak, M. J. Raftery, D. Backhouse, and L. Pereg-Gerk, "Analysis of cotton (*Gossypium hirsutum*) root proteomes during a compatible interaction with the black root rot fungus *Thielaviopsis basicola*," *Proteomics*, vol. 9, no. 2, pp. 335–349, 2009.

[248] H. O. W. O. N. Jung, C. W. O. O. Lim, S. C. Lee, H. W. O. O. Choi, C. H. Hwang, and B. K. Hwang, "Distinct roles of the pepper hypersensitive induced reaction protein gene CaHIR1 in disease and osmotic stress, as determined by comparative transcriptome and proteome analyses," *Planta*, vol. 227, no. 2, pp. 409–425, 2008.

[249] N. Sharma, N. Hotte, M. H. Rahman, M. Mohammadi, M. K. Deyholos, and N. N. V. Kav, "Towards identifying Brassica proteins involved in mediating resistance to *Leptosphaeria maculans*: a proteomics-based approach," *Proteomics*, vol. 8, no. 17, pp. 3516–3535, 2008.

[250] B. Subramanian, V. K. Bansal, and N. N. V. Kav, "Proteome-level investigation of Brassica carinata-derived resistance to *Leptosphaeria maculans*," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 2, pp. 313–324, 2005.

[251] H. Konishi, K. Ishiguro, and S. Komatsu, "A proteomics approach towards understanding blast fungus infection of rice grown under different levels of nitrogen fertilization," *Proteomics*, vol. 1, no. 9, pp. 1162–1171, 2001.

[252] S. U. N. T. A. E. Kim, K. Y. U. S. Cho, S. Yu, et al., "Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells," *Proteomics*, vol. 3, no. 12, pp. 2368–2378, 2003.

[253] S. T. Kim, S. G. Kim, D. H. Hwang, et al., "Proteomic analysis of pathogen-responsive proteins from rice leaves induced by rice blast fungus, *Magnaporthe grisea*," *Proteomics*, vol. 4, no. 11, pp. 3569–3578, 2004.

[254] S. T. Kim, S. Yu, S. G. Kim, et al., "Proteome analysis of rice blast fungus (*Magnaporthe grisea*) proteome during appressorium formation," *Proteomics*, vol. 4, no. 11, pp. 3579–3587, 2004.

[255] M. Liao, Y. Li, and Z. Wang, "Identification of elicitor-responsive proteins in rice leaves by a proteomic approach," *Proteomics*, vol. 9, no. 10, pp. 2809–2819, 2009.

[256] K. U. N. Yuan, B. O. Zhang, Y. Zhang, Q. Cheng, M. Wang, and M. Huang, "Identification of differentially expressed proteins in poplar leaves induced by *Marssonina brunnea* f. sp. Multigermtubi," *Journal of Genetics and Genomics*, vol. 35, no. 1, pp. 49–60, 2008.

[257] C. P. Pirovani, H. A. S. Carvalho, R. C. R. Machado, et al., "Protein extraction for proteome analysis from cacao leaves and meristems, organs infected by *Moniliophthora perniciosa*, the causal agent of the witches' broom disease," *Electrophoresis*, vol. 29, no. 11, pp. 2391–2401, 2008.

[258] F. Wen, H. D. VanEtten, G. Tsaprailis, and M. C. Hawes, "Extracellular proteins in pea root tip and border cell exudates," *Plant Physiology*, vol. 143, no. 2, pp. 773–783, 2007.

[259] Z. Chan, G. Qin, X. Xu, B. Li, and S. Tian, "Proteome approach to characterize proteins induced by antagonist yeast and salicylic acid in peach fruit," *Journal of Proteome Research*, vol. 6, no. 5, pp. 1677–1688, 2007.

[260] M. A. Islam, R. N. Sturrock, and A. K. M. Ekramoddoullah, "A proteomics approach to identify proteins differentially expressed in Douglas-fir seedlings infected by Phellinus sulphurascens," *Journal of Proteomics*, vol. 71, no. 4, pp. 425–438, 2008.

[261] R. C. Amey, T. Schleicher, J. Slinn, et al., "Proteomic analysis of a compatible interaction between Pisum sativum (pea) and the downy mildew pathogen *Peronospora viciae*," *European Journal of Plant Pathology*, vol. 122, no. 1, pp. 41–55, 2008.

[262] T. Cao, S. Srivastava, M. H. Rahman, et al., "Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection," *Plant Science*, vol. 174, no. 1, pp. 97–115, 2008.

[263] J. Lee, T. M. Bricker, M. Lefevre, S. R. M. Pinson, and J. H. Oard, "Proteomic and genetic approaches to identifying defence-related proteins in rice challenged with the fungal pathogen *Rhizoctonia solani*," *Molecular Plant Pathology*, vol. 7, no. 5, pp. 405–416, 2006.

[264] J. Lee, J. Feng, K. B. Campbell, et al., "Quantitative proteomic analysis of bean plants infected by a virulent and avirulent obligate rust fungus," *Molecular and Cellular Proteomics*, vol. 8, no. 1, pp. 19–31, 2009.

[265] Y. U. Liang, S. Srivastava, M. H. Rahman, S. E. Strelkov, and N. N. V. Kav, "Proteome changes in leaves of *Brassica napus* L. as a result of *Sclerotinia sclerotiorum* challenge," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 6, pp. 1963–1976, 2008.

[266] V. Bhadauria, S. Banniza, L.-X. Wang, Y.-D. Wei, and Y.-L. Peng, "Proteomic studies of phytopathogenic fungi, oomycetes and their interactions with hosts," *European Journal of Plant Pathology*, vol. 126, no. 1, pp. 81–95, 2010.

[267] U. Mathesius, "Comparative proteomic studies of root-microbe interactions," *Journal of Proteomics*, vol. 72, no. 3, pp. 353–366, 2009.

[268] A. Mehta, A. C. M. Brasileiro, D. S. L. Souza, et al., "Plant-pathogen interactions: what is proteomics telling us?" *FEBS Journal*, vol. 275, no. 15, pp. 3731–3746, 2008.

[269] Y. U. Liang, H. U. I. Chen, M. Tang, and S. Shen, "Proteome analysis of an ectomycorrhizal fungus *Boletus edulis* under salt shock," *Mycological Research*, vol. 111, no. 8, pp. 939–946, 2007.

[270] A. M. Murad, E. F. Noronha, R. N. Miller, et al., "Proteomic analysis of *Metarhizium anisopliae* secretion in the presence of the insect pest *Callosobruchus maculatus*," *Microbiology*, vol. 154, no. 12, pp. 3766–3774, 2008.

[271] W. U. N. Y. Lin, J. U. I. Y. Chang, C. H. Hish, and T. Z. U. M. Pan, "Profiling the *Monascus pilosus* proteome during nitrogen limitation," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 2, pp. 433–441, 2008.

[272] A. V. Wymelenberg, G. Sabat, D. Martinez, et al., "The *Phanerochaete chrysosporium* secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium," *Journal of Biotechnology*, vol. 118, no. 1, pp. 17–34, 2005.

[273] R. Marra, P. Ambrosino, V. Carbone, et al., "Study of the three-way interaction between *Trichoderma atroviride*, plant and fungal pathogens by using a proteomic approach," *Current Genetics*, vol. 50, no. 5, pp. 307–321, 2006.

[274] S. C. Tseng, S. H. U. Y. Liu, H. H. Yang, C. T. Lo, and K. O. U. C. Peng, "Proteomic study of biocontrol mechanisms of *Trichoderma harzianum* ETS 323 in response to *Rhizoctonia solani*," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 16, pp. 6914–6922, 2008.

[275] V. Seidl, I. S. Druzhinina, and C. P. Kubicek, "A screening system for carbon sources enhancing β-*N*-acetylglucosaminidase formation in *Hypocrea atroviridis* (*Trichoderma atroviride*)," *Microbiology*, vol. 152, no. 7, pp. 2003–2012, 2006.

[276] S. Nagendran, H. E. Hallen-Adams, J. M. Paper, N. Aslam, and J. D. Walton, "Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*," *Fungal Genetics and Biology*, vol. 46, no. 5, pp. 427–435, 2009.

[277] K. J. Welham, M. A. Domin, K. Johnson, L. Jones, and D. S. Ashton, "Characterization of fungal spores by laser desorption/ionization time-of-flight mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 14, no. 5, pp. 307–310, 2000.

[278] M. Both, M. Csukai, M. P. H. Stumpf, and P. D. Spanu, "Gene expression profiles of *Blumeria graminis* indicate dynamic changes to primary metabolism during development of an obligate biotrophic pathogen," *Plant Cell*, vol. 17, no. 7, pp. 2107–2122, 2005.

[279] M. Both, S. E. Eckert, M. Csukai, E. Muller, G. Dimopoulos, and P. D. Spanu, "Transcript profiles of *Blumeria graminis* development during infection reveal a cluster of genes that are potential virulence determinants," *Molecular Plant-Microbe Interactions*, vol. 18, no. 2, pp. 125–133, 2005.

[280] M. R. Remm, C. E. V. Storm, and E. L. L. Sonnhammer, "Automatic clustering of orthologs and in-paralogs from pairwise species comparisons," *Journal of Molecular Biology*, vol. 314, no. 5, pp. 1041–1052, 2001.

[281] J. Grinyer, L. Kautto, M. Traini, et al., "Proteome mapping of the *Trichoderma reesei* 20S proteasome," *Current Genetics*, vol. 51, no. 2, pp. 79–88, 2007.

[282] P. S. Solomon, K. A. R. C. Tan, P. Sanchez, R. M. Cooper, and R. P. Oliver, "The disruption of a Gα subunit sheds new light on the pathogenicity of *Stagonospora nodorum* on wheat," *Molecular Plant-Microbe Interactions*, vol. 17, no. 5, pp. 456–466, 2004.

[283] J. C. Misas-Villamil and R. A. van der Hoorn, "Enzyme-inhibitor interactions at the plant-pathogen interface," *Current Opinion in Plant Biology*, vol. 11, no. 4, pp. 380–388, 2008.

[284] P. Kankanala, K. Czymmek, and B. Valent, "Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus," *Plant Cell*, vol. 19, no. 2, pp. 706–724, 2007.

[285] H. Tjalsma, A. Bolhuis, J. D. H. Jongbloed, S. Bron, and J. M. van Dijl, "Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome," *Microbiology and Molecular Biology Reviews*, vol. 64, no. 3, pp. 515–547, 2000.

[286] R. P. De Vries, "Regulation of *Aspergillus* genes encoding plant cell wall polysaccharide-degrading enzymes; relevance for industrial production," *Applied Microbiology and Biotechnology*, vol. 61, no. 1, pp. 10–20, 2003.

[287] F. M. Freimoser, S. Screen, G. Hu, and R. St. Leger, "EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during growth on insect cuticle," *Microbiology*, vol. 149, no. 7, pp. 1893–1900, 2003.

[288] J. D. Walton, "Deconstructing the cell wall," *Plant Physiology*, vol. 104, no. 4, pp. 1113–1118, 1994.

[289] N. Brito, J. J. Espino, and C. González, "The endo-β-1,4-xylanase Xyn11A is required for virulence in *Botrytis cinerea*," *Molecular Plant-Microbe Interactions*, vol. 19, no. 1, pp. 25–32, 2006.

[290] H. Deising, R. L. Nicholson, M. Haug, R. J. Howard, and K. Mendgen, "Adhesion pad formation and the involvement of cutinase and esterases in the attachment of uredospores to the host cuticle," *Plant Cell*, vol. 4, no. 9, pp. 1101–1111, 1992.

[291] A. Isshiki, K. Akimitsu, M. Yamamoto, and H. Yamamoto, "Endopolygalacturonase is essential for citrus black rot caused by Alternaria citri but not brown spot caused by Alternaria alternata," *Molecular Plant-Microbe Interactions*, vol. 14, no. 6, pp. 749–757, 2001.

[292] B. Oeser, P. M. Heidrich, U. Müller, P. Tudzynski, and K. B. Tenberge, "Polygalacturonase is a pathogenicity factor in the *Claviceps purpurea*/rye interaction," *Fungal Genetics and Biology*, vol. 36, no. 3, pp. 176–186, 2002.

[293] A. ten Have, W. Mulder, J. Visser, and J. A. L. van Kan, "The endopolygalacturonase gene Bcpg1 is required to full virulence of *Botrytis cinerea*," *Molecular Plant-Microbe Interactions*, vol. 11, no. 10, pp. 1009–1016, 1998.

[294] C. A. Voigt, W. Schäfer, and S. Salomon, "A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals," *Plant Journal*, vol. 42, no. 3, pp. 364–375, 2005.

[295] N. Yakoby, D. Beno-Moualem, N. T. Keen, A. Dinoor, O. Pines, and D. Prusky, "Colletotrichum gloeosporioides pelB is an important virulence factor in avocado fruit-fungus interaction," *Molecular Plant-Microbe Interactions*, vol. 14, no. 8, pp. 988–995, 2001.

[296] J. W. Bennett, "The molds of Katrina. Update," *New York Academy of Sciences Magazine*, pp. 6–9, 2006.

[297] G. L. F. Wallis, R. J. Swift, F. W. Hemming, A. P. J. Trinci, and J. F. Peberdy, "Glucoamylase overexpression and secretion in *Aspergillus* niger: analysis of glycosylation," *Biochimica et Biophysica Acta*, vol. 1472, no. 3, pp. 576–586, 1999.

[298] D. Lim, P. Hains, B. Walsh, P. Bergquist, and H. Nevalainen, "Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach," *Proteomics*, vol. 1, no. 7, pp. 899–910, 2001.

[299] M. B. Suárez, L. Sanz, M. I. Chamorro, et al., "Proteomic analysis of secreted proteins from *Trichoderma harzianum*: identification of a fungal cell wall-induced aspartic protease," *Fungal Genetics and Biology*, vol. 42, no. 11, pp. 924–934, 2005.

[300] C. Tian, W. T. Beeson, A. T. Iavarone, et al., "Systems analysis of plant cell wall degradation by the model filamentous fungus Neurospora crassa," *Proceedings of the National*

Academy of Sciences of the United States of America, vol. 106, no. 52, pp. 22157–22162, 2009.

[301] J. M. Cork and M. D. Purugganan, "The evolution of molecular genetic pathways and networks," *BioEssays*, vol. 26, no. 5, pp. 479–484, 2004.

[302] T. Köcher and G. Superti-Furga, "Mass spectrometry-based functional proteomics: from molecular machines to protein networks," *Nature Methods*, vol. 4, no. 10, pp. 807–815, 2007.

[303] W. F. Loomis and P. W. Sternberg, "Genetic networks," *Science*, vol. 269, no. 5224, p. 649, 1995.

[304] M. E. Cusick, N. Klitgord, M. Vidal, and D. E. Hill, "Interactome: gateway into systems biology," *Human Molecular Genetics*, vol. 14, no. 2, pp. R171–R181, 2005.

[305] L. Kiemer and G. Cesareni, "Comparative interactomics: comparing apples and pears?" *Trends in Biotechnology*, vol. 25, no. 10, pp. 448–454, 2007.

[306] L. Giot, J. S. Bader, C. Brouwer, et al., "A protein interaction map of Drosophila melanogaster," *Science*, vol. 302, no. 5651, pp. 1727–1736, 2003.

[307] D. J. LaCount, M. Vignali, R. Chettier, et al., "A protein interaction network of the malaria parasite *Plasmodium falciparum*," *Nature*, vol. 438, no. 7064, pp. 103–107, 2005.

[308] S. Li, C. M. Armstrong, N. Bertin, et al., "A map of the interactome network of the metazoan *C. elegans*," *Science*, vol. 303, no. 5657, pp. 540–543, 2004.

[309] J. R. Parrish, J. Yu, G. Liu, et al., "A proteome-wide protein interaction map for *Campylobacter jejuni*," *Genome Biology*, vol. 8, no. 7, p. R130, 2007.

[310] P. Uetz, L. Glot, G. Cagney, et al., "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.

[311] T. Berggård, S. Linse, and P. James, "Methods for the detection and analysis of protein-protein interactions," *Proteomics*, vol. 7, no. 16, pp. 2833–2842, 2007.

[312] J. A. N. A. Miernyk and J. A. Y. J. Thelen, "Biochemical approaches for discovering protein-protein interactions," *Plant Journal*, vol. 53, no. 4, pp. 597–609, 2008.

[313] B. Causier and B. Davies, "Analysing protein-protein interactions with the yeast two-hybrid system," *Plant Molecular Biology*, vol. 50, no. 6, pp. 855–870, 2002.

[314] L. R. Matthews, P. Vaglio, J. Reboul, et al., "Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs"," *Genome Research*, vol. 11, no. 12, pp. 2120–2126, 2001.

[315] T. Ideker, O. Ozier, B. Schwikowski, and A. F. Siegel, "Discovering regulatory and signalling circuits in molecular interaction networks," *Bioinformatics*, vol. 18, supplement 1, pp. S233–S240, 2002.

[316] X. Wu, L. E. I. Zhu, J. I. E. Guo, D. A. Y. Zhang, and K. U. I. Lin, "Prediction of yeast protein-protein interaction network: insights from the Gene Ontology and annotations," *Nucleic Acids Research*, vol. 34, no. 7, pp. 2137–2150, 2006.

[317] S. E. E. K. Ng, Z. Zhang, and S. H. Tan, "Integrative approach for computationally inferring protein domain interactions," *Bioinformatics*, vol. 19, no. 8, pp. 923–929, 2003.

[318] R. Jothi, M. G. Kann, and T. M. Przytycka, "Predicting protein-protein interaction by searching evolutionary tree automorphism space," *Bioinformatics*, vol. 21, supplement 1, pp. i241–i250, 2005.

[319] U. Ogmen, O. Keskin, A. S. Aytuna, R. Nussinov, and A. Gursoy, "PRISM: protein interactions by structural matching," *Nucleic Acids Research*, vol. 33, no. 2, pp. W331–W336, 2005.

[320] Y. Qi, J. Klein-Seetharaman, and Z. I. V. Bar-Joseph, "A mixture of feature experts approach for protein-protein interaction prediction," *BMC Bioinformatics*, vol. 8, supplement 10, p. S6, 2007.

[321] J. Shen, J. Zhang, X. Luo, et al., "Predicting protein-protein interactions based only on sequences information," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 11, pp. 4337–4341, 2007.

[322] C. von Mering, L. J. Jensen, M. Kuhn, et al., "STRING 7—recent developments in the integration and prediction of protein interactions," *Nucleic Acids Research*, vol. 35, no. 1, pp. D358–D362, 2007.

[323] D. R. Rhodes, S. A. Tomlins, S. Varambally, et al., "Probabilistic model of the human protein-protein interaction network," *Nature Biotechnology*, vol. 23, no. 8, pp. 951–959, 2005.

[324] E. Hirsh and R. Sharan, "Identification of conserved protein complexes based on a model of protein network evolution," *Bioinformatics*, vol. 23, no. 2, pp. e170–e176, 2007.

[325] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold, "Correlation between protein and mRNA abundance in yeast," *Molecular and Cellular Biology*, vol. 19, no. 3, pp. 1720–1730, 1999.

Methodology Report

Differential Proteome Analysis of the Preeclamptic Placenta Using Optimized Protein Extraction

Magnus Centlow,[1] Stefan R. Hansson,[1] and Charlotte Welinder[2]

[1] Department of Obstetrics & Gynecology, Lund University, BMC C14, 221 84 Lund, Sweden
[2] Department of Oncology, Lund University, Barngatan 2B, 221 85 Lund, Sweden

Correspondence should be addressed to Charlotte Welinder, charlotte.welinder@med.lu.se

The human placenta is a difficult tissue to work with using proteomic technology since it contains large amounts of lipids and glycogen. Both lipids and glycogen are known to interfere with the first step in the two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), the isoelectric focusing. In order to gain the best possible protein separation on 2D-PAGE, an optimized sample preparation protocol for placental proteins was developed. Two different buffers, urea/CHAPS and Hepes, were used for solubilization in combination with six different precipitation methods. The removal of glycogen from the samples by centrifugation was crucial for the final proteome maps. Solubilization with urea/CHAPS in combination with dichloromethane/methanol or acidified acetone proved to be the best precipitation procedures. When applied to clinical placenta samples apolipoprotein A1 was found to be accumulated in the preeclamptic placenta, where it may either have a nutritional effect or act as a modifier of signal transduction.

1. Introduction

The human placenta is unique among species. As a matter of fact, of all mammalian organs the placenta shows the greatest variation in terms of anatomy. However, in all species it has the same fundamental function, to nourish the growing fetus, by establishing contact with the maternal blood circulation [1]. In addition, the placenta acts as a protective barrier; as an example it prevents stress hormones to pass over to the fetus via active transport [2]. In order to prevent rejection of the fetus, the placenta expresses an intricate pattern of major histocompatibility complex molecules, immunizing the mother against the foreign fetal tissue [3]. During pregnancy the placenta also has several important endocrine functions, producing important hormones such as human chorionic gonadotropin and human placental lactogen [4, 5]. Hence, due to its vast number of functions the placenta expresses more than 20 000 DNA sequences and is perhaps the organ expressing the largest number of genes [6]. In fact, it has been used as a "fishing pond" for

finding and sequencing new genes and has been used to identify and clone transporters and receptors, such as human growth hormone-variants, integrin alpha V, and subtypes of the angiotensin receptors [7–9].

Preeclampsia (PE) is a pregnancy associated disorder characterized by hypertension and proteinuria. It affects 3%–7% of all pregnancies [10] and is a major cause of maternal and fetal mortality. Treatment for PE is symptomatic, and removal of the placenta is the only curative treatment, suggesting that the placenta is a key in the development of PE. The etiology of PE is still largely unknown; however, it is generally accepted that PE begins with inadequate placentation. Invasion of placental trophoblast cells into the maternal spiral arteries is too shallow, leading to inadequate placental perfusion and hypoxia [11, 12].

Proteomics is a research field undergoing rapid development, especially in regards to the methods being used. The most common method is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). 2D-PAGE for proteome studies gives a map of proteins which reflect changes in

protein expression levels, as well different isoforms and posttranslational modifications. Lately, three types of protein microarrays have become available, namely, analytical microarrays, functional microarrays, and reverse phase microarrays (reviewed in [13]). Highly sensitive and specific recombinant antibody microarrays allowing screening of crude protein extracts are also being developed [14]. Due to the large number of proteins, no protein array platform has yet been designed that covers the entire human placental proteome. Thus, when screening the placental proteome 2D-PAGE is still among the most accepted methods giving the highest coverage although this may soon change. However, there are several obstacles when working with placental tissue, since it contains large amounts of lipids and glycogen [15]. In order for the first dimension, isoelectric focusing (IEF), to be successful, proteins have to be efficiently extracted, and solubilized, and contaminants, such as lipids, polysaccharides, nucleic acids, and salt ions, must be removed. Lipids, polysaccharides and nucleic acids bind to proteins which increases the viscosity of the samples and often results in horizontal streaking on the 2D-PAGE. If the salt concentration is too high in the sample, the conductivity in the IPG strips increases and prolongs the IEF, resulting in poor 2D-PAGE resolution. These contaminants can be removed by precipitation of the solubilized samples. Unfortunately, there is no universal precipitation method that removes all the different contaminants. For example, precipitating samples with dichloromethane/methanol has been described to remove lipids whereas precipitation with ethanol will remove nucleic acids [16]. For these reasons, the placenta proteome is not fully investigated although some progress has been made in the use of proteomics technology to study PE placenta [17, 18].

The primary aim of this study was to establish a robust sample preparation protocol of human placental tissue, by comparing two different solubilization solutions in combination with six different precipitation methods.

2. Material and Methods

2.1. Tissue Samples. Sixty women admitted at the Department of Obstetrics and Gynecology, Lund University Hospital, were included and assigned to two groups: PE ($n = 30$) and control ($n = 30$) (Table 1). All women gave their written informed consent. PE was defined as blood pressure above 140/90 mmHg and proteinuria above 0.3 g/L or rise in blood pressure above 20 mmHg from the first trimester of pregnancy [19]. No women in the control group suffered PE in a previous pregnancy. Patients were matched based on gestational age. A 10 mm cube of placenta tissue was collected immediately after removal of the placenta and was frozen on dry ice and stored at $-80°C$. Patients with other systemic diseases were excluded from the study. This study was approved by the Ethical Committee Review Board for studies in human subjects at Lund University.

2.2. Protein Extraction. Human placental tissues were pulverized, and the frozen placental powder (25 mg frozen weight) was homogenized with a lysis solution containing either 3 mL

TABLE 1: Clinical characteristics of the patients participating in this study. Age and Gestational length are expressed as median (range), and the other values as mean ± standard deviation.

	Controls	PE
n	30	30
Age (years)	32 (23–44)	30 (22–45)
Gestational length (days)	269 (247–295)	265 (235–287)
Systolic MAP (mmHg)	116 ± 11.3	150 ± 12.1
Diastolic MAP (mmHg)	67 ± 4.4	103 ± 7.9
Albumin (g/l)	ND	1.9 ± 1.4
Placenta weight (g)	687 ± 144	631 ± 128
Child gender (M : F)	19 : 11	15 : 15
Child weight (g)	3671 ± 297	3219 ± 571[a]

ND: not detected, M : male, F : female.
[a]Student two-tailed unpaired *t*-test showed a significant difference between the control and the PE group: $P < .0002$.

8 M urea and 2% CHAPS or 3 mL ice cold 10 mM Hepes, pH 7.5, 24 mM KCl supplemented with a protease inhibitor cocktail (Complete mini plus EDTA, Roche, Basel, Switzerland) [20]. Samples were stirred for 1 hour followed by centrifugation at 43 000 g for 2 hours. Protein concentrations were determined by the bicinchoninic acid (BCA) method (Thermo Scientific, Rockford, USA) with bovine serum albumin (BSA) as a standard. Pooled samples were prepared by combining equal amounts of proteins from individual samples. The pooled samples were either used directly (fresh samples) or stored at $-80°C$ prior to use (frozen samples). The individual samples were stored at $-80°C$.

2.3. Protein Precipitation Procedures. Proteins were precipitated according to the following procedures. For 1D-PAGE, 30 μg were precipitated, and for small and large 2D-PAGE, 100 μg and 500 μg were precipitated, respectively.

Acetone. Extracted samples were mixed with ice-cold acetone to final concentration of 80% acetone. Samples were incubated overnight at $-20°C$ followed by centrifugation at 9000 g for 2 minutes. The supernatant was removed, and the protein pellets were allowed to air dry.

Acidified Acetone. Samples were precipitated as above, but the acetone was acidified with acetic acid to a final concentration of 0.05%.

Ethanol. Ten volumes of ice-cold ethanol were added to one volume of sample. Samples were centrifuged as above, and the protein pellets were allowed to air-dry.

Dichloromethane/Methanol. Samples were precipitated as previously described [16]. Briefly, one volume of sample was mixed with 4 volumes of methanol, 2 volumes of dichloromethane, and 3 volumes of distilled water. The samples were mixed between each addition. The samples were centrifuged 9000 g for 2 minutes at room temperature (RT). The supernatant was discarded, and 3 volumes of methanol were added followed by mixing. The samples were

centrifuged at 9000 g for 2 minutes and the supernatant was again discarded after which pellets were air dried.

Trichloroacetic Acid (TCA). One volume of 40% TCA was added to one volume of samples and incubated for 1 hour on ice followed by centrifugation at 9000 g for 2 minutes. The supernatant was removed, and the protein pellets were either directly dried or washed with ethanol and then allowed to air dry. The dried pellets were either dissolved in 30 μL sodium dodecyl sulphate (SDS) sample buffer pH 8.5 (247 mM Tris, 2% SDS, 0.5 mM edetate (EDTA), 16.5 mM dithiothreitol (DTT), 10% glycerol, and 0.025% (w/v) bromophenol blue) followed by 1D-PAGE or dissolved in rehydration sample buffer containing 8 M Urea, 2% CHAPS, 10 mM DTT, and 2% 3–10 NL ampholines (GE Healthcare, Little Chalfont, UK).

2.4. 1D-PAGE. Samples (30 μg) were loaded onto 10-well 4%–12% NuPAGE Bis-Tris gels (Invitrogen, Carlsbad, USA) and run at 200 V for 45 minutes using 2-(N-morpholino) ethanesulfonic acid (MES) as running buffer. Proteins were visualized by staining with Coomassie Brilliant Blue (R-350).

2.5. 2D-PAGE. The first dimensional isoelectric focusing (IEF) was performed on a horizontal Multiphor (Amersham Biosciences, USA). Immobilized pH gradient (IPG) Blue native strips 3–10 NL ($0.5 \times 3 \times 70$ mm or $0.5 \times 3 \times 240$ mm) (SERVA Electrophoresis GmbH, Heidelberg, Germany) were in-gel rehydrated in rehydration sample buffer over night. For 7 cm strips, 130 μL (containing 100 μg of proteins) and for 24 cm strips, 450 μL (containing 500 μg of proteins) rehydration sample buffer were used. The focusing was performed at the following voltage gradients: 150 V for 30 minutes, 300 V for 30 minutes, 600 V for 30 minutes, 1500 V for 1.5 hours, 3000 V for 2 hours (7 cm strips) and 25 hours (24 cm strips). The strips were equilibrated for 10 minutes in a solution containing 65 mM DTT, 6 M urea, 30% (w/v) glycerol, 2% (w/v) SDS and 50 mM Tris-HCl pH 8.8. A second equilibration step was performed for 10 minutes in the same solution except for DTT, which was replaced by 259 mM iodoacetamide. The strips were soaked in electrophoresis buffer (24 mM Tris base, 0.2 M Glycine and 0.1% SDS) just before the second dimensional gel electrophoresis. The strips were applied on 12% homogeneous Tris-glycine gels ($0.1 \times 8 \times 8$ cm or $0.1 \times 20 \times 24$ cm) and overlaid with a solution of 1% agarose in electrophoresis buffer (kept at 60°C). Electrophoresis was carried out in an XCell Mini-cell apparatus (Invitrogen) using electrophoresis buffer at 125 V for 1.5 hours or in an Hoefer DALT gel apparatus (GE Healthcare) at 20°C and constant 80 V for 19 hours.

Gel Staining. Gels were either stained with silver [21] or stained with SYPRO Ruby (Thermo Scientific).

Image Analysis. Small gels were scanned using a CanoScan 995OF (Canon, Solna, Sweden), and large gels were scanned using an Amersham Typhoon 9400 (GE Healthcare). The images of the large gels stained with SYPRO Ruby were submitted to Ludesi 2D Analysis (Ludesi, Malmo, Sweden, http://www.ludesi.com) for spot detection, matching, and analysis.

Peptide Mass Fingerprinting. The spots of interest were washed and digested as previously described [22]. The purified peptides were eluted directly onto the sample target (Anchorchip target, Bruker Daltonik GmbH, Bremen, Germany) where 0.7 μL of matrix, 2,5-dihydroxybenzoic acid (10 mg/ml in 30% ACN) had been added. Mass spectra of positively charged ions were recorded on a Bruker Reflex III instrument (Bruker Daltonik) operated in the reflector mode. A total of 160–210 single shot spectra were accumulated from each sample. The XMASS 5.0 and MS Biotools software packages provided by the manufactures were used for data processing. Known autoproteolysis products from the trypsin were used for internal calibration.

Database Searching. For protein identification, human protein sequences in the SwissProt database were searched using the Mascot Software (Matrix Science Ltd., London, UK). Parameters specified in the search were taxa, *Homo Sapiens*; missed cleavages, 1; peptide mass tolerance +/0.1 Da; fixed modification, carbamidomethyl (C); variable modification, oxidized methionine.

2.6. Western Blot. Proteins were separated by 1D-PAGE using 10-well 12% NuPAGE Bis-Tris gels (Invitrogen) according to manufacturer's instruction. The gels were loaded with 2.5 μg protein per well and run for 60 minutes at 200 V in MOPS running buffer under reduced conditions and blotted onto polyvinylidene fluoride (PVDF) membranes for 8 minutes at 23 V using an iBlot dry blotting system (Invitrogen). Membranes were incubated with primary antibody for 1 hour (mouse monoclonal anti-APOA1, sc-69755, 1 : 2000 dilution), followed by washes and then, incubated with the secondary antibody for 1 hour (goat polyclonal antimouse IgG, sc-2031, 1 : 5000 dilution). Antibodies were obtained from Santa Cruz (Santa Cruz Biotechnology, Santa Cruz, USA). Subsequently, the membrane was exposed to Enhanced chemilumeniscence ECL+ (GE Healthcare) for 3 minutes. Autoradiographic film Hyperfilm ECL (GE Healthcare) was applied to the blot for various durations to obtain satisfactory exposure. When developed, films were scanned with ultra violet light using a Gene Flash scanner (Syngene, Cambridge, UK) and imported into Syngene Gene Tools. One reference sample was added to all gels to serve as an internal calibrator to normalize the quantification between gels. A students t-test was used to determine statistical difference between the two groups. A P-value < .05 was considered statistically significant.

2.7. Immunohistochemistry. Immunohistochemistry was performed using EnVision+System-HRP (DakoCytomation, Carpinteria, USA) on 4 representative PE samples and 4 controls according to the manufacturer's instructions. Briefly, fresh frozen sections, 12 μm thick, of the placenta

(a)　　　　　　　　　(b)　　　　　　　　　(c)
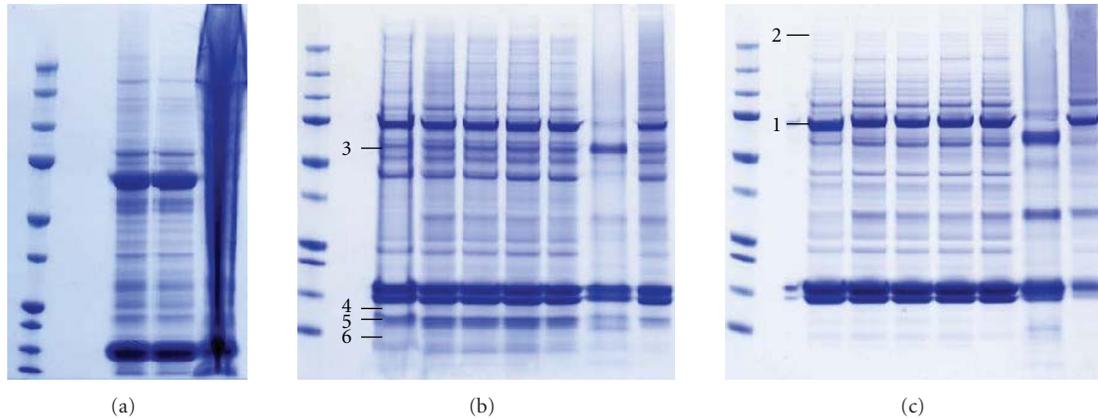
Figure 1: 1D-PAGE comparing solubilization and precipitation methods. (a) One-dimensional separation of proteins in PE placenta solubilized in Hepes buffer before and after centrifugation at 43 000 g. From left to right: Mw Precision standard marker, starting material (30 μg), supernatant after centrifugation (30 μg), and pellet after centrifugation. The gel was stained with Coomassie Brilliant blue. (b) One-dimensional separation of proteins in PE placenta solubilized in urea/CHAPS solution and (c) Hepes buffer. The lanes are from left to right: MW precision standard markers, starting material, and precipitations with acetone, acidified acetone, ethanol, dichloromethane/methanol, TCA, and TCA followed by ethanol wash. Bands indicated with numbers were identified by MALDI-TOF MS (Table 2).

Table 2: Proteins identified using MALDI-TOF MS.

	Protein	Acc. No	Mw (kDa)	Matched peaks	Score	Seq. Cov (%)
Band 1	Annexin A1	P04083	38.9	18	117	59
	Annexin A2	P07355	38.8	20	92	45
	Glyceraldehyd-3-phosphate dehydrogenase	P04406	36.2	13	59	41
Band 2	Endoplasmin	P14625	92.7	15	99	19
	Keratin, type I cytoskeletal 18	P05783	48.0	22	142	37
Band 3	Keratin, type II cytoskeletal 8	P05787	53.7	15	77	26
	Vimentin	P08670	53.7	28	207	52
Band 4	Hemoglobin subunit gamma-1	P69891	16.2	8	86	42
	Hemoglobin subunit gamma-2	P69892	16.2	9	81	51
	Hemoglobin subunit alpha	P69905	15.3	6	66	53
Band 5	Hemoglobin subunit gamma-1	P69891	16.2	8	78	46
	Hemoglobin subunit gamma-2	P69892	16.2	8	78	46
Band 6	Hemoglobin subunit alpha	P69905	15.3	8	72	59
Spot PE	Apolipoprotein A1	P02647	30.7	7	76	26
Spot Ctl	Tropomyosin alpha-3 chain	P06753	32.9	9	63	26

samples were fixed by immersion in 4% phosphate buffered saline (PBS) buffered formaldehyde for 15 minutes at RT. Sections were incubated in a blocking solution containing 3% H_2O_2 for 10 minutes at RT. The APOA1 antibody was diluted (0.1 μg/mL) in PBS containing 0.25% Triton X100 and 1.5% normal goat serum. Sections were incubated over night at 4°C after which the sections were rinsed. Peroxidase labeled polymer was applied to cover the specimens and then incubated for 30 minutes. After rinsing the sections, LiquidDAB substrate chromogen solution was applied and incubated for 8 minutes. The sections were rinsed and stained with Mayers hematoxylin after which sections were mounted with Mountex (Histolab, Gothenburg, Sweden) and cover slipped. Sections were viewed under an Olympus BX-60 microscope and images captured using an Olympus

DP50CU digital camera (Olympus America Inc., Center Valley, USA).

3. Results

3.1. Solubilization and Precipitation. Two solubilization solutions were chosen: Hepes buffer [20] and the commonly used urea/CHAPS. A centrifugation step was added to remove contaminants such as glycogen and cell debris. The pellet resulted in smearing on 1D-PAGE (Figure 1(a)) and without the centrifugation step; no proteins entered the IEF-strip resulting in a blank 2D-PAGE (data not shown). There were no obvious differences in protein patterns between the supernatant and the starting material (Figure 1(a)). 1D-PAGE separation was used to determine protein expression
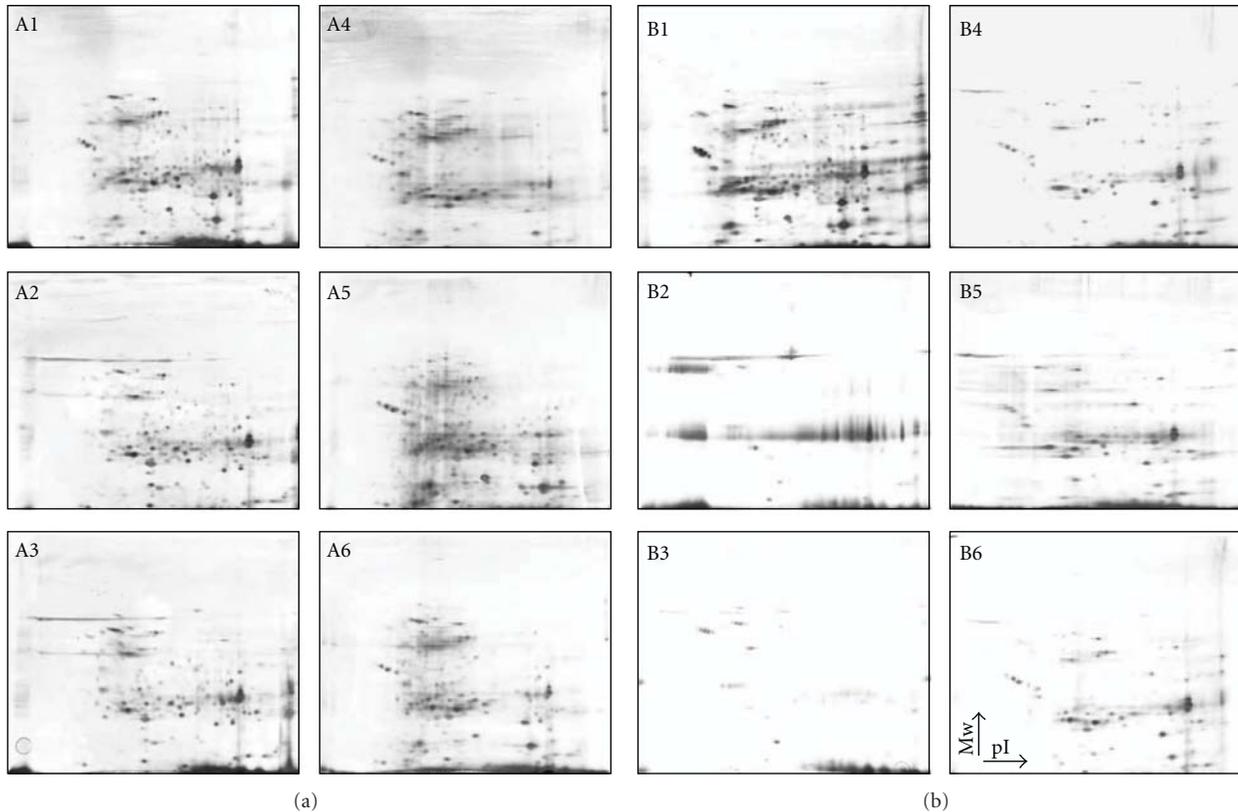
FIGURE 2: Two-dimensional separation of precipitated proteins (100 μg) in PE placenta solubilized in (a) urea/CHAPS and (b) Hepes buffer. Precipitation procedures (1) dichloromethane/methanol, (2) TCA, (3) TCA followed by ethanol wash, (4) acetone, (5) acidified acetone, and (6) ethanol. The gels were stained with silver. Arrows on lower right image represent directions for increasing Mw and pI. Directions are the same for all gels.

from the two different solubilization solutions. A major difference in the protein pattern between Hepes buffer and urea/CHAPS was obtained in the low molecular mass range. Minor differences at higher molecular mass range are also obtained when comparing the two different solubilization solutions (Figures 1(b) and 1(c)). A number of the protein bands that visibly differed between the two solubilization solutions were identified using MALDI-TOF MS. The obtained identities included annexin A1, annexin A2, endoplasmin, vimentin, and hemoglobin subunit alpha and gamma (Table 2). Some of the bands from the 1D-PAGE contained more than one identity due to the limited power of resolution using 1D-PAGE.

Six different precipitation procedures were tested: acetone, acidified acetone, ethanol dichloromethane/methanol, TCA, and TCA followed by ethanol wash. To determine the effects of precipitation procedures on protein patterns, precipitated samples were analyzed on 1D-PAGE gels (Figures 1(b) and 1(c)). Samples precipitated with TCA, with or without ethanol wash, showed a different protein expression profile compared to the other four precipitation methods. No significant differences in the 1D-PAGE expression profiles were detected between the other four precipitation procedures (acetone, acidified acetone, ethanol and dichloromethane/methanol).

3.2. 2D-PAGE. Samples solubilized with urea/CHAPS generally resulted in more protein spots and less horizontal streaking, regardless of precipitation procedure (Figure 2). The investigation of whether freezing of solubilized and centrifuged samples influenced the results, both fresh and frozen samples, was analyzed using 2D-PAGE. No major differences could be detected in the protein expression profiles for any of the different solubilization or precipitation methods (data not shown). An increased number of low molecular weight proteins were obtained with samples solubilized in urea/CHAPS precipitated with acidified acetone (Figure 2, A5).

These spots were not obtained with other combinations of solubilizations and precipitations (Figure 3). Electroendosmotic flow (EOF) was detected during the IEF run with samples precipitated with acetone, acidified acetone, or ethanol. This resulted in swelling of the IPG-strip in the alkaline pH gradient and might cause a poor protein separation at an isoelectric point above 8.

3.3. Preeclampsia. Urea/CHAPS in combination with either acidified acetone or dichloromethane/methanol precipitation was judged to be the most efficient extraction methods for 2D-PAGE. However, we chose dicholoromethane/methanol since it is known to remove lipids of which
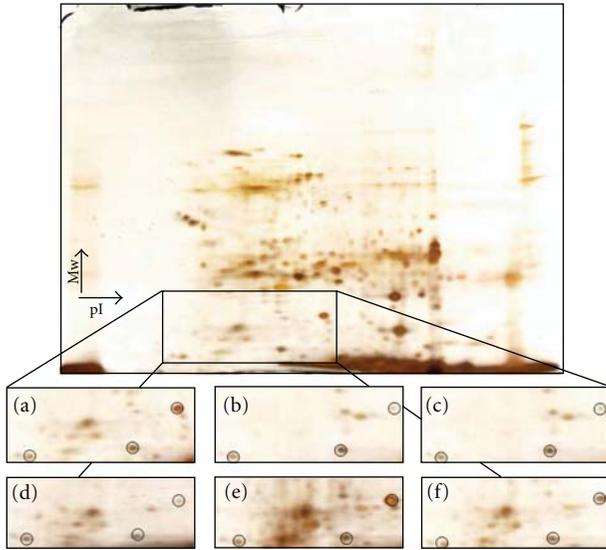
FIGURE 3: Two-dimensional separation of precipitated proteins (100 μg) in PE placenta solubilized in urea/CHAPS. Segments show differential expression pattern in consequence of precipitation procedures. (a) dichloromethane/methanol, (b) TCA, (c) TCA followed by ethanol wash, (d) acetone, (e) acidified acetone, and (f) ethanol. Circles indicate landmarks for orientation. The gels were stained with silver.

there is abundance in the placenta. In a first attempt to investigate whether it is possible to detect differences between the PE and normal placenta in crude extracts, we pooled crude placenta extracts from 30 preeclamptic and 30 normotensive women respectively. In order to decrease the variation of the differently expressed proteins on the 2D-PAGE, three replicates of each pooled group were made. The number of spots detected was 604 and 747 for preeclamptic and normotensive, respectively. The number of spots was calculated based on the criteria that the spots were present in all triplicate gels in each group. The analysis of the 2D-PAGE showed that 23 proteins were increased and 28 were decreased in the PE placenta (Figure 4). Two proteins have so far been identified with MALDI-TOF MS: apolipoprotein A1 (APOA1) (acc number: P02647) and tropomyosin-3 (TPM3) (acc number: P06753). APOA1 was 1.63 times increased in the PE placenta pool as compared to the control pool (PE: 10871 ± 1340 optical density (OD) units and controls: 6663 ± 1772 OD units values presented as mean $\pm$ SD). TPM3 was only found in the normotensive pool.

The increased accumulation of APOA1 in the PE placenta was validated by means of western blot (Figure 5(a)). All samples, 30 PE and 30 controls, were run individually. After measuring the absorption of all the APOA1 bands, APOA1 was found to be 1.18 times increased in PE (PE: 63120 ± 14350 optical density (OD) units and controls: 52810 ± 13310 OD units values presented as mean $\pm$ SD). The accumulation of APOA1 was significantly increased in PE ($P = 0.02$) (Figure 5(b)). Immunohistochemistry showed APOA1 to be accumulated in trophoblasts. Both normotensive and

PE placenta sections showed a similar pattern with weaker staining in the control sections (Figures 5(c)–5(e)). The TPM3 expression could not be validated by means of Western blot since the recommended antibodies did not work in our hands (sc-32516 and sc-18174, Santa Cruz Biotechnology, Santa Cruz, USA).

4. Discussion

Due to the vast number of functions and many different cell types besides the trophoblasts, including white and red blood cells and endothelial cells, the placenta is a complex organ to study. In order to optimize the protein sample preparation for 2D-PAGE, we compared two different solubilization solutions in combination with six different precipitation methods in order to obtain maximal separation of the proteins.

Hepes buffer and urea/CHAPS showed different protein patterns in the low molecular mass range. Although no protease inhibitors were added to the urea/CHAPS solution, urea denatures and inactivates proteases and thus, it is unlikely that the differences are due proteolytic cleavage [23]. Several steps in the sample preparations appear to be important to remove contaminating substances. During pregnancy there is a gradual accumulation of glycogen as well as a large amount of lipids in the placenta [15]. Glycogen and lipids are known to disrupt IEF-separation by blocking the IPG-strip resulting in no or very few spots on 2D-PAGE. Hence, the centrifugation step was crucial for the sample preparation. The centrifugation did not cause a selective loss of proteins, indicating that most of the placental proteins remained in the supernatant. The freezing of samples directly after solubilization (without the centrifugation step) resulted in insoluble aggregates. However, freezing samples after the centrifugation step did not change the proteome map. This observation may facilitate the work flow in proteomics where many samples are analyzed, and the samples can be stored in the freezer until further use without having a major influence on the results.

Our optimization procedures showed that solubilizing with urea and CHAPS and precipitating with dichloromethane/methanol or acidified acetone gave the highest number of spots on a 2D-PAGE. Solubilization with Hepes buffer in combination with dichloromethane/methanol also gave a high number of spots (Figure 2, B1), although there was considerably more horizontal streaking. Dichloromethane/methanol has the advantage or also removing lipids, and since the placenta is rich in lipids, this precipitation method may be the preferred choice. Hence, in a first study urea/CHAPS in combination with dichloromethane/methanol was chosen to examine the differences between the PE and normotensive placental proteome.

The differentiating proteins have proven to be hard to identify due to the low concentration of the proteins in the differing spots. The same problem was highlighted in a recent preeclamptic proteomic study, where, as a consequence of low amount of proteins, none of the differentially expressed proteins could be identified [18]. However, we have so far been able to identify two proteins differing between PE and

(a)

(b)
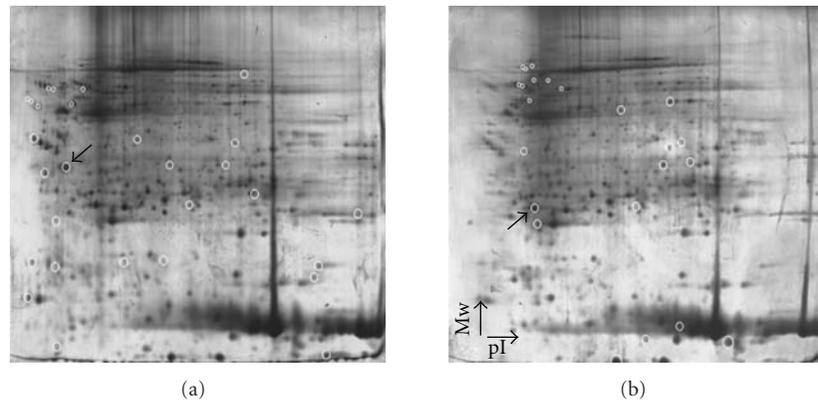
FIGURE 4: Two-dimensional separation of pooled proteins from normotensive (a) and PE (b). Circles indicate spots whose amount was increased in the 2D-PAGE of normotensive and PE samples analyzed with the Ludesi's software Redfin. Arrows show identified proteins; see Table 2. The gels were stained with SYPRO Ruby.

(a)

(b)

(c)

(d)

(e)
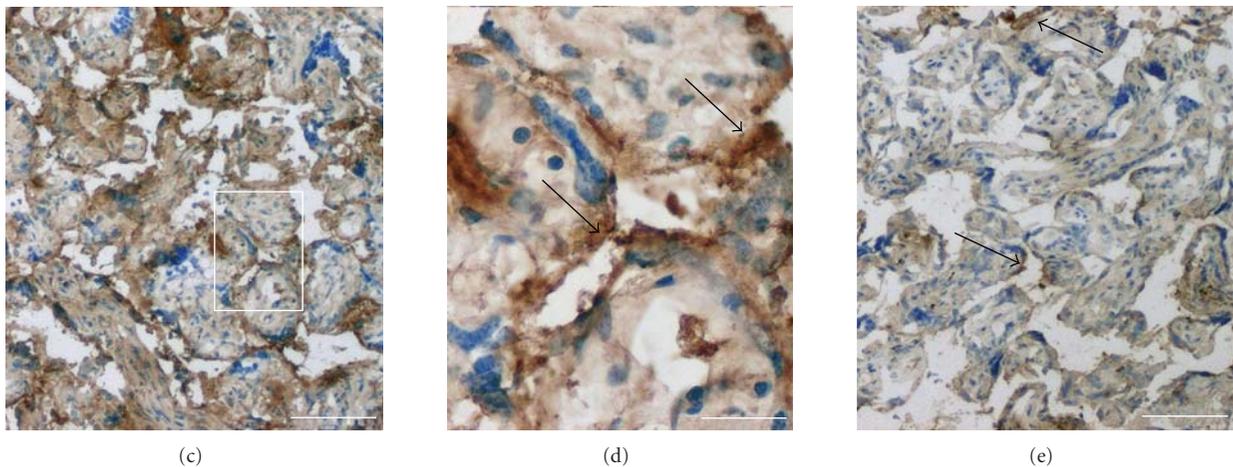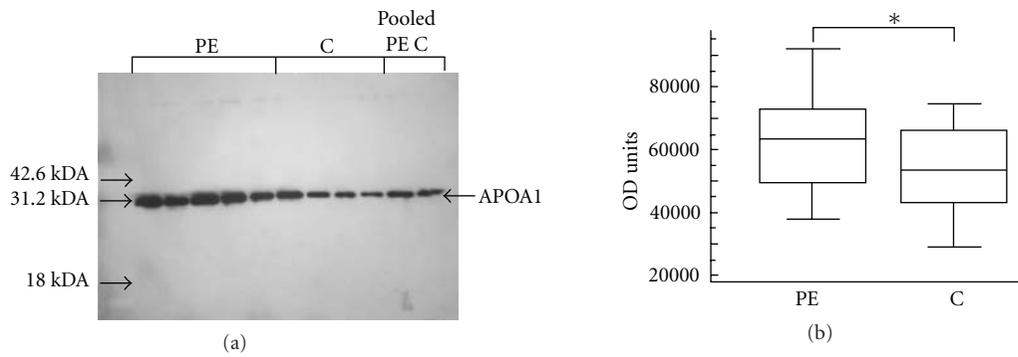
FIGURE 5: Validation of APOA1 accumulation. (a) A representative scanned image of Western blots for APOA1 (PE: $n = 5$ and C: $n = 4$). All individual sample extracts were run (PE: $n = 30$ and C: $n = 30$). Expected size for APOA1 is 28 kDA. Molecular weight markers are represented by arrows. (b) Box plot for the intensity data for APOA1 as determined by GeneTools, showing the median and the 25th and 75th percentile. APOA1 was significantly increased in PE ($P = .02$) (PE: 62247 (56597–67897) OD units, C: 53551 (48354–58749) OD units. Both values are presented as mean (95th percentile confidence interval). (c) Representative image for immunohistochemistry on APOA1 on a PE section. Scale bar = 100 μm. (d) Zoomed in image of APOA1 staining in a PE section (from the box in D). APOA1 staining was detected in trophoblasts (arrow). Scale bar = 30 μm. (e) A representative image of a control section. APOA1 staining was weaker than in PE but was localized in trophoblasts as seen in PE. Scale bar = 100 μm.

controls: APOA1 (increased in PE) and TPM3 (decreased in PE Figure 4). The increase of APOA1 in PE was validated by Western blot and immunohistochemistry (Figure 5). Due to the difficulties in identifying low abundance proteins in a crude extracted sample, one might have to do purifications of specific cell types, for example, trophoblast to explore the proteome of the placenta [24]. Another limitation with 2D-PAGE is that membrane proteins do not easily enter IEF, the first dimension since they often precipitate in the IPG-strip.

In our first attempt to study differences between PE and normotensive placenta, we found apolipoprotein A1, a lipid binding protein, to be accumulated in the PE placenta. Of the lipoproteins, APOA1 is mainly synthesized in the liver and intestines and is the major fraction of the high-density lipoproteins [25]. The placenta has also been suggested as a site of synthesis for APOA1 although it is not clear whether this is an actual de novo synthesis or an accumulation from the blood stream. In a recent whole genome microarray analysis, increased APOA1 gene expression was not detected in the PE placenta. It is unlikely that the accumulation of APOA1 in the PE placenta is due to high levels of APOA1 in the maternal plasma since plasma levels APOA1 are not increased in PE [26, 27]. In a recent proteomics study of amniotic fluid, accumulation of pro-APOA1 was shown in the amniotic fluid in PE [28]. Hence, amniotic fluid may be a possible source of the accumulated APOA1 seen in the PE placenta. How the accumulation of APOA1 affects the PE placenta is unknown. APOA1 may server either as nutrient or as a signaling molecule by affecting biological signaling pathways such as protein kinase C dependent pathway (PKC and the mitogen-activated protein kinases (MAPK) [29, 30].

In conclusion, two solubilization methods in combination with six different precipitation procedures were tested. Several steps in the sample preparation appear important. Removal of glycogen by centrifugation is crucial since the glycogen clogged the IPG-strip. Both the solubilization solution and the precipitation procedure affected the placenta proteome maps. In this study, solubilization with urea/CHAPS in combination with dichloromethane/methanol or acidified acetone proved to be the best precipitation procedures, resulting in more spots and less streaking on 2D-PAGE. Importantly, freezing samples following centrifugation of the solubilized samples did not affect the proteome map. However, due to the limitation using crude extracts for 2D-PAGE, low abundance proteins cannot readily be detected, and thus subcellular fractioning may be required to fully investigate the placenta proteome.

Acknowledgments

References

[1] H. N. Jones, T. L. Powell, and T. Jansson, "Regulation of placental nutrient transport—a review," *Placenta*, vol. 28, no. 8-9, pp. 763–774, 2007.

[2] V. Ganapathy, P. D. Prasad, and F. H. Leibach, "Use of human placenta in studies of monoamine transporters," *Methods in Enzymology*, vol. 296, pp. 278–290, 1998.

[3] J. S. Hunt, "Stranger in a strange land," *Immunological Reviews*, vol. 213, no. 1, pp. 36–47, 2006.

[4] J. C. Cross, "Placental function in development and disease," *Reproduction, Fertility and Development*, vol. 18, no. 2, pp. 71–76, 2006.

[5] D. T. Krieger, "Placenta as a source of 'brain' and 'pituitary' hormones," *Biology of Reproduction*, vol. 26, no. 1, pp. 55–71, 1982.

[6] G. Blaich, B. Janssen, G. Roth, and J. Salfeld, "Overview: differentiating issues in the development of macromolecules compared with small molecules," in *Handbook of Pharmaceutical Biotechnology*, S. C. Gad, Ed., p. 77, John Wiley & Sons, New York, NY, USA, 1st edition, 2007.

[7] S. A. Liebhaber, M. Urbanek, J. Ray, R. S. Tuan, and N. E. Cooke, "Characterization and histologic localization of human growth hormone-variant gene expression in the placenta," *The Journal of Clinical Investigation*, vol. 83, no. 6, pp. 1985–1991, 1989.

[8] J. W. Smith, D. J. Vestal, S. V. Irwin, T. A. Burke, and D. A. Cheresh, "Purification and functional characterization of integrin $\alpha(v)\beta5$. An adhesion receptor for vitronectin," *The Journal of Biological Chemistry*, vol. 265, no. 19, pp. 11008–11013, 1990.

[9] H. Konishi, S. Kuroda, Y. Inada, and Y. Fujisawa, "Novel subtype of human angiotensin II type 1 receptor: cDNA cloning and expression," *Biochemical and Biophysical Research Communications*, vol. 199, no. 2, pp. 467–474, 1994.

[10] J. M. Roberts and D. W. Cooper, "Pathogenesis and genetics of pre-eclampsia," *The Lancet*, vol. 357, no. 9249, pp. 53–56, 2001.

[11] I. A. Brosens, W. B. Robertson, and H. G. Dixon, "The role of the spiral arteries in the pathogenesis of preeclampsia," *Obstetrics and Gynecology Annual*, vol. 1, pp. 177–191, 1972.

[12] N. Soleymanlou, I. Jurisica, O. Nevo, et al., "Molecular evidence of placental hypoxia in preeclampsia," *The Journal of Clinical Endocrinology & Metabolism*, vol. 90, no. 7, pp. 4299–4308, 2005.

[13] D. A. Hall, J. Ptacek, and M. Snyder, "Protein microarray technology," *Mechanisms of Ageing and Development*, vol. 128, no. 1, pp. 161–167, 2007.

[14] C. Wingren and C. A. Borrebaeck, "Antibody-based microarrays," *Methods in Molecular Biology*, vol. 509, pp. 57–84, 2009.

[15] N. M. Gude, C. T. Roberts, B. Kalionis, and R. G. King, "Growth and function of the normal human placenta," *Thrombosis Research*, vol. 114, no. 5-6, pp. 397–407, 2004.

[16] D. Wessel and U. I. Flugge, "A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids," *Analytical Biochemistry*, vol. 138, no. 1, pp. 141–143, 1984.

[17] R. P. Webster and L. Myatt, "Elucidation of the molecular mechanisms of preeclampsia using proteomic technologies," *Proteomics: Clinical Applications*, vol. 1, no. 9, pp. 1147–1155, 2007.

[18] K. Mine, A. Katayama, T. Matsumura, et al., "Proteome analysis of human placentae: pre-eclampsia versus normal pregnancy," *Placenta*, vol. 28, no. 7, pp. 676–687, 2007.

[19] F. Milne, C. Redman, J. Walker, et al., "The pre-eclampsia community guideline (PRECOG): how to screen for and detect onset of pre-eclampsia in the community," *British Medical Journal*, vol. 330, no. 7491, pp. 576–580, 2005.

[20] R. Hass and C. Sohn, "Increased oxidative stress in pre-eclamptic placenta is associated with altered proteasome activity and protein patterns," *Placenta*, vol. 24, no. 10, pp. 979–984, 2003.

[21] A. Shevchenko, M. Wilm, O. Vorm, and M. Mann, "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels," *Analytical Chemistry*, vol. 68, no. 5, pp. 850–858, 1996.

[22] T. Berggård, S. Miron, P. Önnerfjord, et al., "Calbindin D_{28k} exhibits properties characteristic of a Ca^{2+} sensor," *The Journal of Biological Chemistry*, vol. 277, no. 19, pp. 16662–16672, 2002.

[23] S. Singh, D. W. Powell, M. J. Rane, et al., "Identification of the p16-Arc subunit of the Arp 2/3 complex as a substrate of MAPK-activated protein kinase 2 by proteomic analysis," *The Journal of Biological Chemistry*, vol. 278, no. 38, pp. 36410–36417, 2003.

[24] J. M. Robinson, D. D. Vandré, and W. E. Ackerman IV, "Placental proteomics: a shortcut to biological insight," *Placenta*, vol. 30, supplement 1, pp. 83–89, 2009.

[25] A. L. Wu and H. G. Windmueller, "Relative contributions by liver and intestine to individual plasma apolipoproteins in the rat," *The Journal of Biological Chemistry*, vol. 254, no. 15, pp. 7316–7322, 1979.

[26] M. B. Cekmen, A. B. Erbagci, A. Balat, et al., "Plasma lipid and lipoprotein concentrations in pregnancy induced hypertension," *Clinical Biochemistry*, vol. 36, no. 7, pp. 575–578, 2003.

[27] K. Winkler, B. Wetzka, M. M. Hoffmann, et al., "Triglyceride-rich lipoproteins are associated with hypertension in preeclampsia," *The Journal of Clinical Endocrinology & Metabolism*, vol. 88, no. 3, pp. 1162–1166, 2003.

[28] J. S. Park, K.-J. Oh, E. R. Norwitz, et al., "Identification of proteomic biomarkers of preeclampsia in amniotic fluid using SELDI-TOF mass spectrometry," *Reproductive Sciences*, vol. 15, no. 5, pp. 457–468, 2008.

[29] J.-M. Darbon, J.-F. Tournier, J.-P. Tauber, and F. Bayard, "Possible role of protein phosphorylation in the mitogenic effect of high density lipoproteins on cultured vascular endothelial cells," *The Journal of Biological Chemistry*, vol. 261, no. 17, pp. 8002–8008, 1986.

[30] Y. Q. Wu, E. V. Jorgensen, and S. Handwerger, "High density lipoproteins stimulate placental lactogen release and adenosine $3',5'$-monophosphate (cAMP) production in human trophoblast cells: evidence for cAMP as a second messenger in human placental lactogen release," *Endocrinology*, vol. 123, no. 4, pp. 1879–1884, 1988.

Methodology Report

A Proteomic Approach for Plasma Biomarker Discovery with iTRAQ Labelling and OFFGEL Fractionation

Emilie Ernoult, Anthony Bourreau, Erick Gamelin, and Catherine Guette

Centre INSERM Régional de Recherche sur le Cancer U892, Centre Régional de Lutte Contre le Cancer Paul Papin,
2 rue Moll, 49933 Angers Cedex 9, France

Correspondence should be addressed to Catherine Guette, c.guette@unimedia.fr

Human blood plasma contains a plethora of proteins, encompassing not only proteins that have plasma-based functionalities, but also possibly every other form of low concentrated human proteins. As it circulates through the tissues, the plasma picks up proteins that are released from their origin due to physiological events such as tissue remodeling and cell death. Specific disease processes or tumors are often characterized by plasma "signatures," which may become obvious via changes in the plasma proteome profile, for example, through over expression of proteins. However, the wide dynamic range of proteins present in plasma makes their analysis very challenging, because high-abundance proteins tend to mask those of lower abundance. In the present study, we used a strategy combining iTRAQ as a reagent which improved the peptide ionization and peptide OFFGEL fractionation that has already been shown, in our previous research, to improve the proteome coverage of cellular extracts. Two prefractioning methods were compared: immunodepletion and a bead-based library of combinatorial hexapeptide technology. Our data suggested that both methods were complementary, with regard to the number of identified proteins. iTRAQ labelling, in association with OFFGEL fractionation, allowed more than 300 different proteins to be characterized from 400 μg of plasma proteins.

1. Introduction

Blood circulates throughout every part of the body, and no other biofluid has the same degree of intimacy with the body. Therefore, it is not surprising that it possesses such a wealth of information concerning the overall pathophysiology of a patient. As an example, alterations in protein abundance can serve to indicate pathological abnormalities: diseases, toxic effects of clinical treatments and so forth.

The choice between plasma and serum has been abundantly documented in the literature [1, 2]. When blood is collected, many changes occur in the proteins it contains, due to the presence of proteolytic enzymes (proteases) and other enzymes, which remain active in the blood sample during handling and processing. The HUPO Committee and its research collaborators concluded with the recommendation that plasma is the preferred specimen taken from the blood. The reasons for this are (i) less ex vivo degradation, and

(ii) much less variability than in the case of the protease-rich process of clotting. Misek et al. [3] showed with Cy5-, Cy3-, Cy2-labeled serum and plasma on DIGE-2D-PAGE, after extensive fractionation of intact proteins before tryptic digestion, that isoforms of abundant proteins were more often shifted to lower-than-expected MW in serum than in plasma. Tammen et al. [4] reported that 40% of the low-MW peptides detected were serum specific.

Biomarker discovery in plasma is often limited by the availability of sufficient volumes. It is also complicated by the wide dynamic range of the human plasma proteome, which comprises proteins spanning concentrations of more than 11 orders of magnitude, with the top 10 most abundant plasma proteins accounting for approximatively 90% of the total plasma proteins. Potential disease biomarkers are often present in low concentrations, and the dynamic range of the plasma proteins poses a significant analytical challenge to proteomic approaches. A prefractionation method is

necessary in the biomarker process. The most common technique is immunodepletion, which has been extensively used for the specific removal of high abundance proteins, based on the action of specific antibodies [5, 6]. More recently, saturation protein binding to a random peptide library has been proposed as an alternative method [7, 8].

One of the methods used to discover biomarkers is the identification and quantification of proteins, based on an iTRAQ quantitative proteomic approach. iTRAQ is ideally suited for biomarker applications, as it provides both quantification and multiplexing in a single reagent, and has been applied to the analysis of clinical samples such as human cerebrospinal fluid, and disease tissues, and has been used for the in vitro profiling of cells to identify differentially expressed proteins. To the best of our knowledge, there are currently only two published papers where iTRAQ has been used to study human serum and plasma. Hergenroeder et al. [9] employed iTRAQ and electrospray ionization tandem mass spectrometry (ESI-MS/MS) to serum depleted of 12 high abundance proteins, leading to the identification of 160 proteins. Song et al. [10] used iTRAQ protocol and MALDI-MS/MS to identify 105 proteins in human plasma.

We recently demonstrated that iTRAQ labelling and peptide OFFGEL fractionation in a first dimension improved the identification of weakly concentrated proteins from a cellular extract [11]. The aim of the present study was firstly, to use iTRAQ as reagent to improve the MALDI ionization of peptides and secondly to evaluate the performance of our previous strategy for the study of the human plasma proteome, in terms of the number of identified proteins, the presence of high abundance proteins, and the identification of medium and weakly concentrated proteins.

2. Materials and Methods

2.1. Human Blood Plasma Samples. A citrated plasma pool, composed of a collection of 10 methylene blue virus-inactivated plasma samples, obtained from healthy donors using apheresis, was provided by the French National Public Blood Institution (Etablissement Français du Sang Bourgogne Franche Comté, CHU Le Bocage, Dijon, France). The plasma pool (Internal Quality Control) was loaded into 0.5 ml bar-coded straws, stored at 4°C for 2 hours, and then transferred into liquid nitrogen. The plasma straws were packed in a dry ice container during transportation to our laboratory and were immediately stored in a −80°C freezer until they were needed.

2.2. Immunoaffinity Depletion of High-Abundance Proteins. In a four independent experiments, the 14 most highly abundant proteins were removed from the plasma, using antibody-based depletion with a Human 14 Multiple Affinity Removal System, MARS-Hu 14 (Agilent Technologies, Santa Clara, CA, USA). Two different spin cartridges were used to deplete 10 times 10 μL of 0.22 μm-filtered plasma. This process required 2 buffers, A and B (Agilent Technologies).

The pH 7.4 phosphate salt-containing buffer A was used for the equilibration, loading and washing steps. Flow-through fractions containing low-abundance proteins were collected and stored at −80°C until they were ready for analysis. A pH 2.5 urea buffer B was used for elution of the bound, highly abundant proteins from the cartridge. The experiment was conducted at room temperature according to the protocol supplied by the manufacturer.

2.3. Hexapeptide Ligand Library Treatment. Plasma proteins were "equalized" using the ProteoMiner Protein Enrichment Kit (Bio-rad laboratories, Hercules, CA, USA). Four different spin columns were run in parallel, according to the manufacturer's instructions. Each was loaded with 900 μL of 0.22 μm-filtered plasma, for 2 hours at room temperature, and with 100 μL of 1 M sodium citrate, and 20 mM of HEPES, at pH 7.4. No bead agglomeration was observed. The proteins were desorbed using a two-step elution. The first beads were incubated twice with 100 μL of the kit elution reagent (4 M urea, 1% (w/v) CHAPS, 5% (v/v) acetic acid) for 15 minutes. Then, 100 μL of 6 M guanidine-HCl, at pH 6.0, were added twice for 15 minutes. For each column, the four elution fractions were pooled and stored at −80°C, until they were needed for analysis.

2.4. Buffer Exchange and Protein Content Estimation. Using 2000 MWCO Hydrosart Vivaspin 2 spin concentrators (Sartorius Stedim Biotech, Göttingen, Germany), the pre-fractionated plasma samples were concentrated and buffer-exchanged, by subjecting them to repeated (four times) centrifugation, to an appropriate 0.5 M triethylammonium bicarbonate (TEAB) pH 8.5 buffer (Sigma-Aldrich Corporation, Saint Louis, MO, USA), for downstream analysis. The protein concentrations of whole and prefractionated plasma samples were determined using a FluoroProfile Protein Quantification Kit (Sigma-Aldrich Corporation), with BSA as the standard.

2.5. 1-Dimensional Gel Electrophoresis (1DGE). Equal amounts of proteins from crude and prefractionated plasma, obtained by immunoaffinity depletion or hexapeptide ligand library treatment, were diluted in an SDS loading buffer at 2 mg/mL, and heated to 100°C for 10 minutes. The elution beads from one ProteoMiner column were then washed and directly boiled at 100°C for 10 minutes in a 100 μL SDS loading buffer. The proteins were then separated onto a home-made 12.5% SDS-PAGE gel (16 cm long and 1.5 mm thick), using a standard Laemmli buffer system in an SE 600 Ruby electrophoresis unit (GE Healthcare, Chalfont Saint Giles, UK). Precision Plus Protein Standards (Bio-rad laboratories) were loaded in the molecular weight marker lane. The gel image was acquired on an Ettan DIGE Imager (GE Healthcare), after total protein fluorescent poststaining with Deep Purple (excitation, 532 nm; emission 610 nm), according to the standard protocol (GE Healthcare).

2.6. Protein Digestion and Peptide iTRAQ Labelling. 100 μg of protein solutions, at 5 mg/mL in 0.5 M TEAB, at pH 8.5, were

4 human blood plasma aliquots

Prefractionation

Protein trypsic digestion

Peptide iTRAQ labelling

114 115 116 117

Pool: 400 μg proteins

Peptide OFFGEL 24 fractionation

Nano RPLC-MALDI MS/MS
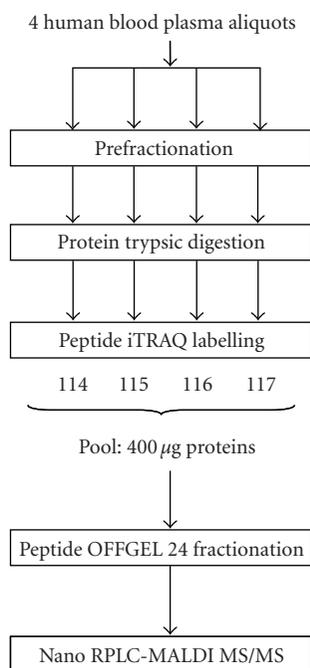
FIGURE 1: Illustration of the workflow.

reduced for 1 hour at 60°C with 5 mM tris-(2-carboxyethyl) phosphine (TCEP) and were cysteine-blocked with 10 mM methyl methanethiosulfonate (MMTS) at room temperature for 10 minutes. The proteins were then digested for 40 hours at 37°C, by 10 μg of TPCK-treated trypsin, with $CaCl_2$ (Applied Biosystems, Foster City, CA, USA). Each peptide solution was labelled for 3 hours at room temperature, using an iTRAQ reagent previously reconstituted in 70 μL of ethanol, according to the iTRAQ Reagents Multiplex Kit protocol (Applied Biosystems). The reaction was stopped by adding milliQ water, and the samples labelled, respectively, with 114, 115, 116, and 117 mass-tagged iTRAQ reagents were combined according to the experimental protocol shown in Figure 1.

2.7. Peptide OFFGEL-IEF Fractionation. For pI-based peptide separation, the 3100 OFFGEL Fractionator (Agilent Technologies) was used with a 24-well setup. Prior to electrofocusing, the peptide samples were desalted onto a Sep-Pak C18 cartridge (Waters Corporation, Milford, MA, USA) and were resolubilized in 3.6 mL of 5% (v/v) glycerol and 1% (v/v) IPG buffer, at pH 3–10 (GE Healthcare). The 24 cm-long IPG gel strips (GE Healthcare), with a 3–10 linear pH range, were rehydrated for 15 minutes according to the manufacturer's manual. Then, 150 μL of sample was loaded in each of the 24 wells.

Electrofocusing of the peptides was performed at 20°C until a level of 50 kVh was reached. After focusing, the 24 peptide fractions were withdrawn and the wells were rinsed with 200 μL of a solution of milliQ water/methanol/formic acid (49/50/1). After 15 minutes, each of the rinsing solutions

was pooled with its corresponding peptide fraction. All of the fractions were evaporated by centrifugation under vacuum and were maintained at −20°C. Just prior to nano-LC separation, the fractions were resuspended in 20 μL of milliQ water with 0.1% (v/v) TFA.

2.8. Nano-LC Separation. The peptide fractions were separated on an Ultimate 3000 nano-LC system (Dionex Corporation, Sunnyvalle, CA, USA), using a C18 column (PepMap100, 3 μm, 100 A, 75 μm id ×15 cm, Dionex Corporation) at a flow rate of 300 nL/minute. Buffer A comprised 2% ACN in milliQ water, with 0.05% TFA, and buffer B comprised 80% ACN in milliQ water, with 0.04% TFA. The peptide solutions were first desalted for 3 minutes using buffer A only on the precolumn, and the separation occurred over a period of 70 minutes, with the following gradient: 0 to 20% B in 10 minutes, 20% to 55% B in 55 minutes, and 55% to 100% B in 5 minutes. Chromatograms were recorded at a wavelength of 214 nm. Following a 12-minute run, the peptide fractions were collected for 10 seconds using a Probot microfraction collector (Dionex Corporation), and spotted directly onto a MALDI sample plate (1664 spots per plate, Applied Biosystems). The CHCA matrix (LaserBioLabs, Sophia-Antipolis, France), with a concentration of 2 mg/mL in 70% ACN, in milliQ water with 0.1% TFA, was continuously added to the column effluent via a micro "T" mixing piece at a flow rate of 1.2 μL/min.

2.9. MALDI-MS/MS. MS and MS/MS analyses of offline spotted peptide samples were performed using the 4800 MALDI-TOF/TOF Analyser (Applied Biosystems). After screening of all LC-MALDI sample positions in MS-positive reflector mode, using 1500 laser shots, the fragmentation of automatically selected precursors was performed at a collision energy of 1 kV using air as the collision gas (pressure ~2 ×10⁻⁶ Torr). MS spectra were acquired between *m/z* 800 and 4000. The parent ion of the Glu-1 fibrinopeptide at *m/z* 1570.677, diluted in the matrix (3 femtomoles per spot), was used for internal calibration. Up to 12 of the most intense ion signals per spot position, characterised by an S/N > 12, were selected as precursors for MS/MS acquisition. Peptide and protein identifications were performed using ProteinPilot Software v 2.0 (Applied Biosystems) and the Paragon algorithm [12]. Each MS/MS spectrum was searched against the Uniprot/Swissprot database (release 96, September 2008) for Homo Sapiens species, with the fixed modification of methyl methanethiosulfonate-labelled cysteine parameter enabled. Other parameters such as the tryptic cleavage specificity, the precursor ion mass accuracy and the fragment ion mass accuracy, are MALDI 4800 built-in functions of the ProteinPilot software. The ProteinPilot software calculated a confidence percentage (the unused score), which reflects the probability of a hit being a "false positive," meaning that at the 95% confidence level, there is a false positive identification probability of about 5%. While this software automatically accepts all peptides with an identification confidence level > 1%, only proteins having at least one peptide above the 95% confidence level were

initially recorded. Low confidence peptides cannot give a positive protein identification by themselves but may support the presence of a protein identified using other peptides with higher confidence. Searches against a concatenated database containing both forward and reversed sequences enabled the false discovery rate to be kept below 1%.

2.10. Data Analysis. In order to analyse the quality of pI fractionation after OFFGEL-IEF and MALDI-MS/MS identification, the experimental pI was calculated for each peptide, using the pI/MW tool of the ExPASy Proteomic Server [13] checking all the deamidation modifications which could influence its value. Then, the average experimental pI of peptides (after filtering for false positive responses) was compared, for each of the 24 fractions, with the theoretical pH values provided by Agilent Technologies for 24 cm-long IPG gel strips with a 3–10 linear pH range. To study the relative abundance of proteins in the plasma, the MS/MS spectra, which enabled protein identification with at least 2 peptides, were counted for each protein [14].

3. Results and Discussion

Our strategy for the study of the human plasma proteome was based on three-step fractionation. In the first step, the plasma samples were prefractionated using either an immunodepletion method, or a peptide ligand library strategy. The proteins were then digested by trypsin, resulted peptides were iTRAQ-labelled and OFFGEL-fractionated in 24 fractions. Each fraction was then analysed by nano-LC on a C18 column (Figure 1).

3.1. Identification of Proteins. The experimental design for the iTRAQ labeling of proteins from the immunodepleted and bead-treated plasma was the same. The prefractionated plasma samples were concentrated and dissolved in the appropriate iTRAQ buffer using spin concentrators before the steps of reduction, MMTS blocking, digestion and iTRAQ labelling (Figure 1). After OFFGEL separation of 400 μg of iTRAQ-labelled peptides in 24 fractions, Protein Pilot software leads to the identification of 332 proteins in immunodepleted plasma and 320 proteins from the hexapeptide ligand library treated plasma (Figure 2).

The average experimental pH value of each OFFGEL fraction is indicated by a bar in Figure 3. The theoretical pH values provided by the manufacturer are also shown by a dashed line. The pI value for each identified peptide was calculated using Bjellqvist's algorithm [15], without taking the iTRAQ groups in the N-term position, and/or the lateral lysine chain, into account. Using these data, average pI values with standard deviations were calculated for all of the peptides identified in each fraction (Figure 3). The average experimental pI value deviated from the theoretical pI value by an average error of ±0.90, for both prefractionation strategies. From the immunodepleted plasma, 243 proteins were characterized by at least 2 peptides, and from the plasma treated by the hexapeptide bead library, 228 were associated
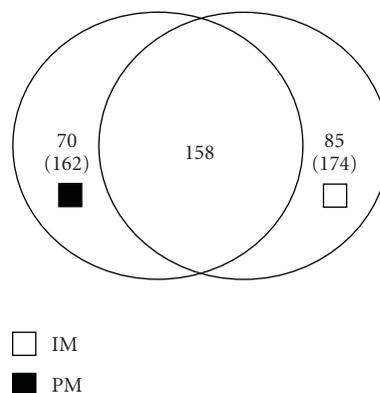
Figure 2: Venn diagram presenting the number of total plasma proteins identified with two or more peptides (in brackets, with one peptide) after immunoaffinity depletion of high-abundant proteins (IM), hexapeptide ligand library treatment (PM) or both prefractionation methods.
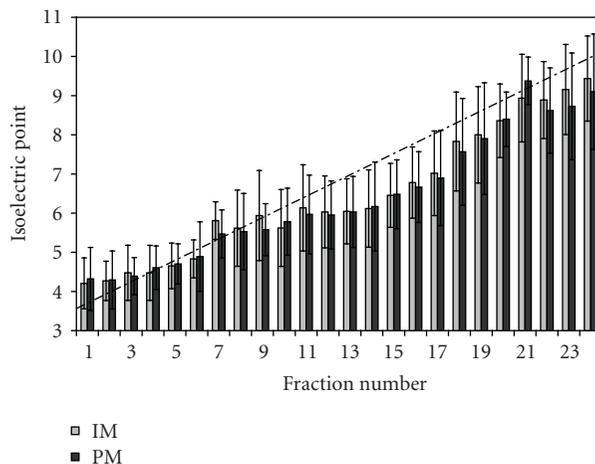
Figure 3: Analysis of the quality of the OFFGEL fractionation of iTRAQ-labelled peptides from immunoaffinity depleted (IM) or hexapeptide ligand library treated (PM) plasma. The average experimental pI of all peptides in each of the 24 OFFGEL fractions after filtering for false positive is presented as bars. Error bars indicate the SD of each fraction's experimental pI. The broken line is based on the theoretical pI values for an IPG strip of 24 cm, pH 3–10 provided by Agilent Technologies.

with at least 2 peptides, suggesting that both prefractioning methods produced virtually the same number of identified proteins.

Among these, 158 were common to both methods (Figure 2). Nevertheless, in addition to these mutual proteins, 85 proteins (with at least 2 peptides) were identified by immunodepletion technology only, and 70 proteins (with at least 2 peptides) were identified by the equalizer strategy only, suggesting that these strategies are complementary. The merging of both sets of data allowed a total of 313 proteins with at least 2 peptides to be identified.

TABLE 1: Protein recovery after plasma prefractionation using immunoaffinity depletion of high-abundant proteins or hexapeptide ligand library treatment. Protein content was estimated using FluoroProfile Protein Quantification Kit.

	MARS-Hu 14				Proteo Miner			
Plasma volume loaded (μL)	10×10 depletion cycles				900			
Protein quantity loaded (mg) (whole plasma : 63 mg/mL proteins)	6,3				56,7			
Protein quantity recovered after prefractionation (μg)	340	401	415	299	1420	1158	1195	1572
Average protein quantity recovered after prefractionation (μg)	364 (or 36.4 μg per depletion cycle)				1336			
% of total protein mass removal by prefractionation	94.2				97.6			

A previous study conducted by us with 400 μg of immunodepleted plasma and treated in the same conditions, except iTRAQ labelling, showed 115 identified proteins (Supplementary Material available online at doi:10.1155/2010/927917) in agreement with Song et al. results [10]. This result demonstrated the efficiency of iTRAQ labelling for the peptide ionization and the protein identification according our previous study [11].

3.2. Evaluation of the First Prefractioning Step. The human plasma was prefractionated using two different methods: an immunodepletion strategy on a human MARS-14, which depleted the 14 most abundant plasma proteins, and a peptide ligand library technology with a ProteoMiner column, which should "equalize" the plasma proteins. The reproducibility of these approaches was evaluated by four independent experiments (Figure 1). The total protein content was used to investigate reproducibility and protein recovery (Table 1). The total protein concentration of untreated plasma was 63 mg/mL. The mean recovery rate of the eluted bead-treated plasma was 2.4% (in agreement with previous results [8]), compared with 5.8% following the depletion process using the MARS-14 column. Both methodologies showed a reproducibility of around 15%, in the determination of total protein content.

For each approach, reproducibility of biological experiment and iTRAQ labelling were evaluated with the coefficient variation calculation from ProteinPilot results. An average of ±18% variation across the 4 experiments with the human MARS-14 strategy and an average of ±11% with the ProteoMiner approach were exhibited. Separation of native plasma, immunodepleted, and bead-treated plasma samples by SDS-PAGE revealed a significant reduction in the dynamic range of protein concentration in the treated fractions, when compared with native plasma (Figure 4) but did not identify one prefractionation method as being superior to the other. Comparison of the 20 most abundant proteins evaluated by the MS/MS spectra counting technique indicated that both prefractioning methods were equivalent, in terms of the estimated protein concentrations (Figure 5). With ProteoMiner treatment, fibrinogen alpha and beta chains were the most commonly found proteins, thus suggesting that this technique could be more suitable for serum samples.

3.3. Gene Ontology Annotations. Gene ontology analysis for cellular localization revealed that a large proportion
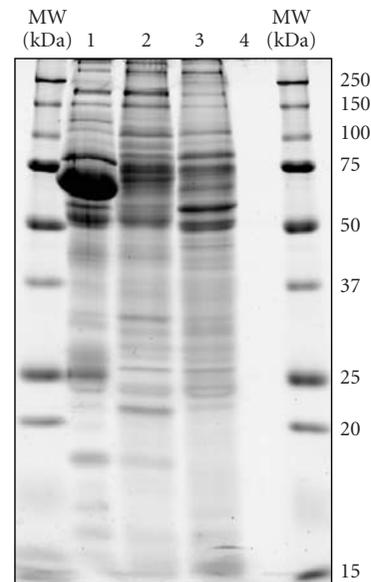
FIGURE 4: 1DGE analysis of human blood plasma before and after prefractionation using immunoaffinity depletion of high-abundant proteins or hexapeptide ligand library treatment. An equal amount of proteins from whole plasma (lane 1), plasma fraction from MARS-Hu 14 (lane 2), plasma fraction from ProteoMiner (lane 3), and potentially remaining proteins from boiled in SDS loading buffer ProteoMiner beads after elution (lane 4) were separated on 12.5% SDS-PAGE gel and were stained with Deep Purple.

of proteins, with predicted extracellular locations (33%) (Figure 6(b)) was present in our plasma proteome map. Functional classification also revealed that most of the proteins are involved in "binding" and enzyme activities (Figure 6(a)).

Among these proteins, we identified medium concentrated proteins with concentrations ranging around 30 ng/mL, such as P-selectin, cadherin 5 [16, 17] (Table 2). We successfully detected the low-concentrated proteins Hepatocyte growth factor activator, insulin like growth factor binding-protein2 and Sex hormone-binding globulin in the both experiments with ProteinPilot unused score >2 (proteins identified with at least 2 peptides; 95% confidence). The literature data showed that the concentration of these proteins was in the range of 8–18 ng/mL [16, 18, 19]. Compared with the concentration of the most abundant
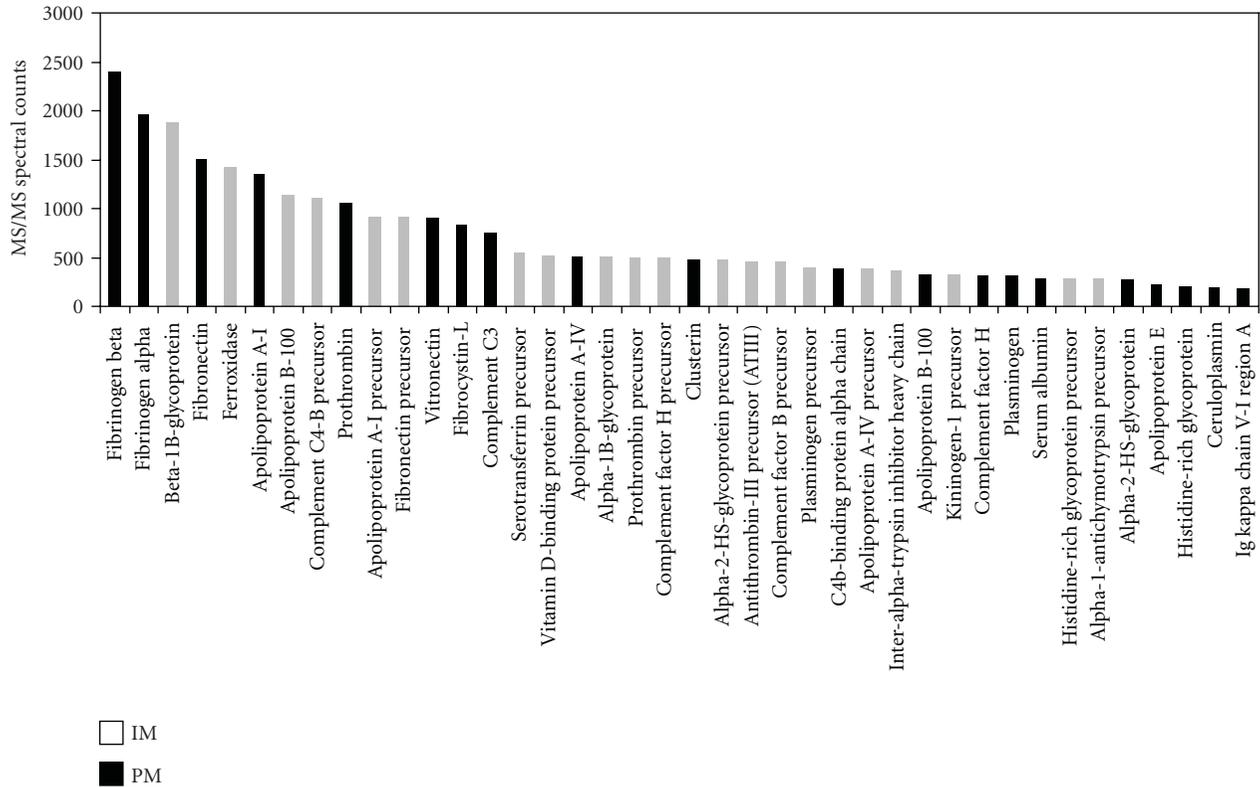
FIGURE 5: Relative abundance of the twenty most abundant proteins determined by counting the number of MS/MS spectra for each protein identified after immunoaffinity depletion (IM) or Hexapeptide ligand library treatment (PM) of plasma.

(a)   (b)
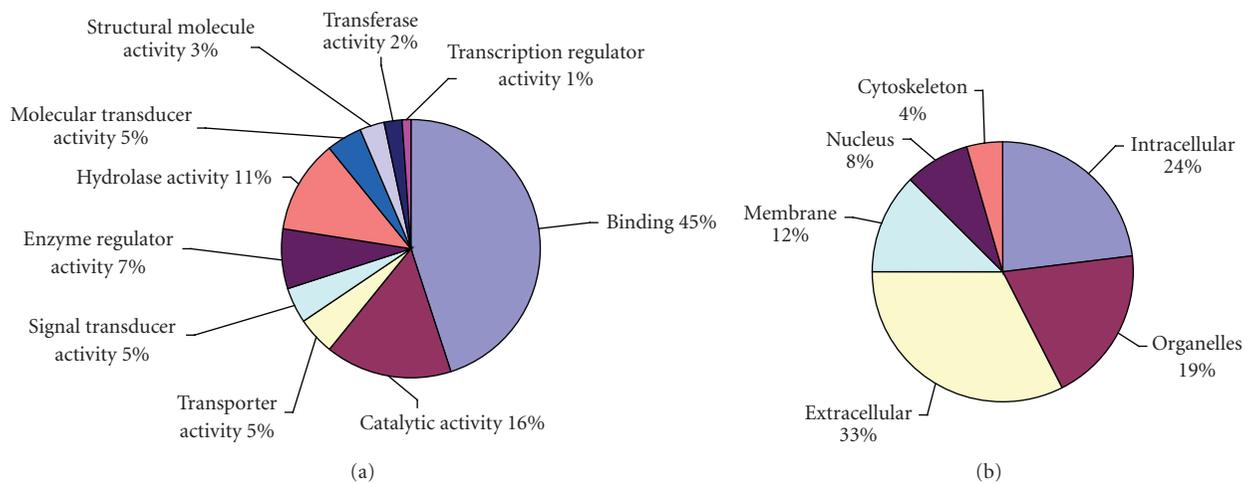
FIGURE 6: Pie charts showing the gene ontology classification of identified plasma proteins according to cellular function (a) and to cellular component category (b).

plasma protein (HSA) which is around 50 mg·mL, we can conclude that the dynamic range to detect low-abundant plasma proteins could be extended to 10^6-10^7.

4. Conclusions

The number of proteins which could be identified in 400 μg of plasma proteins was markedly increased in this study,

when compared to similar samples studied without iTRAQ labelling, or with iTRAQ labelling, but without OFFGEL fractionation [10], suggesting that our strategy improved the proteome coverage of human plasma.

The limited number of individual proteins identified in this study, despite prefractionation, highlights the challenge of plasma-based biomarker discovery. From our experience, similar iTRAQ analyses of cellular extracts are able to identify

TABLE 2: Some of the weakly abundant plasma proteins identified with two or more peptides after immunoaffinity depletion of high-abundant proteins (IM) or hexapeptide ligand library treatment (PM). Associated concentrations were found in the literature in the mid-range from 5 to 50 000 ng/mL serum or plasma.

Protein name	Prefractionation	Concentration	Bibliographic reference
Plasma retinol-binding protein	PM / IM	44.4 μg/mL (plasma)	[20]
Kallistatin	IM	22.1 μg/mL (plasma)	[21]
Ficolin-3	IM	21.6 μg/mL (plasma)	[22]
Tetranectin	IM	13.75 μg/mL (serum)	[23]
Selenoprotein P	IM	5.3 μg/mL (plasma)	[24]
Von Willebrand factor	PM / IM	1.3 μg/mL (plasma)	[16]
Intelectin-1	PM	0.1 to 1.0 μg/mL (serum)	[25]
Extracellular SOD	PM / IM	150 ng/mL (plasma)	[26]
Mannose-binding lectin (Protein C)	IM	97 ng/mL (plasma)	[16]
Insulin-like growth factor binding protein 3	PM / IM	59 ng/mL (plasma)	[16]
P-selectin	IM	30 ng/mL (plasma)	[17]
Cadherin 5	PM / IM	30 ng/mL (plasma)	[16]
Macrophage colony-stimulating factor 1 receptor (M-CSF R)	IM	26 ng/mL (plasma)	[16]
Hepatocyte growth factor activator	PM / IM	17.6 ng/mL (serum)	[18]
Insulin-like growth factor binding protein 2	PM / IM	15 ng/mL (plasma)	[16]
Sex hormone-binding globulin	PM / IM	to 8.1 ng/mL (plasma)	[19]

more than 1000 different proteins [11]. Clearly, the presence of many highly abundant proteins in human plasma and therefore, after trypsin digestion, the presence of many highly concentrated peptides prevent a good MALDI ionization of weak-concentrated peptides and therefore limit the depth of analysis. Theses results argue in favour of the search for new strategies for the removal of abundant plasma proteins [27], or for the enrichment of less abundant proteins, in order to facilitate the efficient discovery of biomarkers.

Acknowledgment

References

[1] A. J. Rai, C. A. Gelfand, B. C. Haywood, et al., "HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples," *Proteomics*, vol. 5, no. 13, pp. 3262–3277, 2005.

[2] G. S. Omenn, D. J. States, M. Adamski, et al., "Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database," *Proteomics*, vol. 5, no. 13, pp. 3226–3245, 2005.

[3] D. E. Misek, R. Kuick, H. Wang, et al., "A wide range of protein isoforms in serum and plasma uncovered by a quantitative intact protein analysis system," *Proteomics*, vol. 5, no. 13, pp. 3343–3352, 2005.

[4] H. Tammen, I. Schulte, R. Hess, et al., "Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display," *Proteomics*, vol. 5, no. 13, pp. 3414–3422, 2005.

[5] N. Zolotarjova, J. Martosella, G. Nicol, J. Bailey, B. E. Boyes, and W. C. Barrett, "Differences among techniques for high-abundant protein depletion," *Proteomics*, vol. 5, no. 13, pp. 3304–3313, 2005.

[6] Y. Y. Wang, P. Cheng, and D. W. Chan, "A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis," *Proteomics*, vol. 3, no. 3, pp. 243–248, 2003.

[7] P. G. Righetti, E. Boschetti, L. Lomas, and A. Citterio, "Protein equalizer technology : the quest for a "democratic proteome"," *Proteomics*, vol. 6, no. 14, pp. 3980–3992, 2006.

[8] C. Sihlbom, I. Kanmert, H. von Bahr, and P. Davidsson, "Evaluation of the combination of bead technology with SELDI-TOF-MS and 2-D DIGE for detection of plasma proteins," *Journal of Proteome Research*, vol. 7, no. 9, pp. 4191–4198, 2008.

[9] G. Hergenroeder, J. B. Redell, A. N. Moore, et al., "Identification of serum biomarkers in brain-injured adults: potential for predicting elevated intracranial pressure," *Journal of Neurotrauma*, vol. 25, no. 2, pp. 79–93, 2008.

[10] X. Song, J. Bandow, J. Sherman, et al., "iTRAQ experimental design for plasma biomarker discovery," *Journal of Proteome Research*, vol. 7, no. 7, pp. 2952–2958, 2008.

[11] E. Ernoult, E. Gamelin, and C. Guette, "Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation," *Proteome Science*, vol. 6, article 27, 2008.

[12] I. V. Shilov, S. L. Seymourt, A. A. Patel, et al., "The Paragon Algorithm, a next generation search engine that uses sequence temperature values sequence temperature values and feature probabilities to identify peptides from tandem mass spectra," *Molecular and Cellular Proteomics*, vol. 6, no. 9, pp. 1638–1655, 2007.

[13] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.

[14] J. M. Asara, H. R. Christofk, L. M. Freimark, and L. C. Cantley, "A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen," *Proteomics*, vol. 8, no. 5, pp. 994–999, 2008.

[15] B. Bjellqvist, G. J. Hughes, Ch. Pasquali, et al., "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences," *Electrophoresis*, vol. 14, no. 10, pp. 1023–1031, 1993.

[16] B. B. Haab, B. H. Geierstanger, G. Michailidis, et al., "Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data," *Proteomics*, vol. 5, no. 13, pp. 3278–3291, 2005.

[17] H. M. Amin, S. Ahmad, J. M. Walenga, D. A. Hoppensteadt, H. Leitz, and J. Fareed, "Soluble P-selectin in human plasma: effect of anticoagulant matrix and its levels in patients with cardiovascular disorders," *Clinical and Applied Thrombosis/Hemostasis*, vol. 6, no. 2, pp. 71–76, 2000.

[18] K. F. Wader, U. M. Fagerli, R. U. Holt, B. Stordal, M. Borset, A. Sundan, and A. Waage, "Elevated serum concentrations of activated hepatocyte growth factor activator in patients with multiple myeloma," *Eur. J. Haematol*, vol. 81, pp. 380–383, 2008.

[19] S. K. Cunningham, T. Loughlin, M. Culliton, and T. J. McKenna, "Plasma sex hormone-binding globulin levels decrease during the second decade of life irrespective of pubertal status," *Journal of Clinical Endocrinology and Metabolism*, vol. 58, no. 5, pp. 915–918, 1984.

[20] F. R. Smith, A. Raz, and D. S. Goodman, "Radioimmunoassay of human plasma retinol-binding protein," *The Journal of Clinical Investigation*, vol. 9, pp. 1754–1761, 1970.

[21] J. Chao, A. Schmaier, L.-M. Chen, Z. Yang, and L. Chao, "Kallistatin, a novel human tissue kallikrein inhibitor: levels in body fluids, blood cells, and tissues in health and disease," *Journal of Laboratory and Clinical Medicine*, vol. 127, no. 6, pp. 612–620, 1996.

[22] C. B. Svendsen, T. Hummelshøj, L. Munthe-Fog, et al., "Ficolins and mannose-binding lectin in danish patients with sarcoidosis," *Respiratory Medicine*, vol. 102, no. 9, pp. 1237–1242, 2008.

[23] E. F. Kamper, A. D. Papahiliss, M. K. Angelopoulou, et al., "Serum levels of tetranectin, intercellular adhesion molecule-1 and interleukin-10 in B-chronic lymphocytic leukemia," *Clinical Biochemistry*, vol. 32, no. 8, pp. 639–645, 1999.

[24] Y. Saito, Y. Watanabe, E. Saito, T. Honjoh, and K. Takahashi, "Production and application of monoclonal antibodies to human selenoprotein P," *J. Health Sci.*, vol. 47, pp. 346–352, 2001.

[25] R.-Z. Yang, M.-J. Lee, H. Hu, et al., "Identification of omentin as a novel depot-specific adipokine in human adipose tissue: possible role in modulating insulin action," *American Journal of Physiology*, vol. 290, pp. E1253–E1261, 2006.

[26] J. Sandström, P. Nilsson, K. Karlsson, and S. L. Marklund, "10-fold increase in human plasma extracellular superoxide dismutase content caused by a mutation in heparin-binding domain," *Journal of Biological Chemistry*, vol. 269, no. 29, pp. 19163–19166, 1994.

[27] Y. Tanaka, H. Akiyama, T. Kuroda, et al., "A novel approach and protocol for discovering extremely low-abundance proteins in serum," *Proteomics*, vol. 6, no. 17, pp. 4845–4855, 2006.

Methodology Report

Enhanced MALDI-TOF MS Analysis of Phosphopeptides Using an Optimized DHAP/DAHC Matrix

Junjie Hou,[1,2] Zhensheng Xie,[1] Peng Xue,[1] Ziyou Cui,[1,2] Xiulan Chen,[1,2] Jing Li,[1,2] Tanxi Cai,[1] Peng Wu,[1] and Fuquan Yang[1]

[1] *Proteomic Platform and National Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China*
[2] *Department of Biology, Graduate University of the Chinese Academy of Sciences, Beijing 100049, China*

Correspondence should be addressed to Fuquan Yang, fqyang@sun5.ibp.ac.cn

Selecting an appropriate matrix solution is one of the most effective means of increasing the ionization efficiency of phosphopeptides in matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). In this study, we systematically assessed matrix combinations of 2, 6-dihydroxyacetophenone (DHAP) and diammonium hydrogen citrate (DAHC), and demonstrated that the low ratio DHAP/DAHC matrix was more effective in enhancing the ionization of phosphopeptides. Low femtomole level of phosphopeptides from the tryptic digests of α-casein and β-casein was readily detected by MALDI-TOF-MS in both positive and negative ion mode without desalination or phosphopeptide enrichment. Compared with the DHB/PA matrix, the optimized DHAP/DAHC matrix yielded superior sample homogeneity and higher phosphopeptide measurement sensitivity, particularly when multiple phosphorylated peptides were assessed. Finally, the DHAP/DAHC matrix was applied to identify phosphorylation sites from α-casein and β-casein and to characterize two phosphorylation sites from the human histone H1 treated with Cyclin-Dependent Kinase-1 (CDK1) by MALDI-TOF/TOF MS.

1. Introduction

Reversible phosphorylation of serine, threonine, and tyrosine residues is one of the most common and important regulatory posttranslational modifications of proteins and plays a crucial role in many biological processes, including cellular signal transduction, regulation of enzyme activities, metabolic regulation, molecular recognition and biomolecular interaction, protein localization, cell differentiation and proliferation [1–5]. Therefore, identifying phosphorylated proteins is an important step to elucidate the structural and functional role of the protein's phosphorylation regulation.

Up to now, mass spectrometry, including matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and electrospray ionization mass spectrometry (ESI-MS) have been key technologies in characterizing protein phosphorylation and phosphoproteome, and each approach presented strengths and limitations [6–15]. However, because of the low and dynamic

stoichiometry of phosphorylation on proteins [16, 17] and the low ionization efficiency of phosphopeptides, it remains challenging to thoroughly characterize phosphorylation sites using mass spectrometry [18]. Considerable effort has been made recently to improve the ionization efficiency of phosphopeptides by selecting and optimizing appropriate matrices with MALDI-TOF-MS. Several materials have been employed in phosphopeptide analysis, including 2, 5-dihydroxybenzoic acid (DHB) [19–21], 2, 4, 6-trihydroxyacetophenone (THAP) [22], and ionic liquid matrices [20]. In addition, ammonium salts, including diammonium hydrogen citrate (DAHC) [22–24], ammonium acetate [23] and monoammonium phosphate [25], and phosphoric acid [19–21] have been used as matrix additives and have shown some ability to enhance phosphopeptide ionization during MALDI-TOF-MS.

2, 6-Dihydroxyacetophenone (DHAP) was first used by Gorman et al. as a matrix for MALDI analysis of fragile peptides, disulfide bonding, and small proteins, including

phosphopeptides [24]. In recent years, Xu et al. reported that the DHAP/DAHC matrix combination was used as an effective MALDI matrix in detecting methyl esterified phosphopeptides by MALDI-TOF-MS [26, 27]. After systematic optimization in this study, we found that contrary to previous reports, the matrix forms at a lower ratio of DHAP/DAHC and that this ratio shows greater efficiency in improving ionization ability of phosphopeptides. We investigated the characteristics of an optimal DHAP/DAHC matrix, and analyzed morphology and detection sensitivity of phosphopeptides. Using tryptic digests of α- and β-casein, we demonstrated a CID MS/MS analysis of phosphopeptides by MALDI-TOF/TOF-MS with an optimized DHAP/DAHC matrix. Finally, we used this approach to analyze phosphorylation sites on human histone H1 treated with Cyclin-Dependent Kinase-1 (CDK1).

2. Experimental

2.1. Materials.
The chemicals 2, 6-dihydroxyacetophenone (DHAP), 2, 5-dihydroxy-benzoic acid (DHB), diammonium hydrogen citrate (DAHC), ammonium bicarbonate, trifluoroacetic acid (TFA), alkaline phosphatases, α-casein and β-casein (both from bovine milk, purity by electrophoresis) were all purchased from Sigma Aldrich (St. Louis, MO, USA). Cyclin-Dependent Kinase-1(CDK1) and human histone H1 were purchased from New England BioLabs, Inc. The standard peptides VNQIG**p**TLSESIK (**p**T, purity > 95%) and SGSLHRI**p**YTHQS (**p**Y, purity > 85%) were synthesized in Beijing Scilight Biotechnology Ltd. Co. (Beijing, China). Sequencing grade modified trypsin (porcine) was purchased from Promega Co. (Madison, WI, USA). HPLC-grade acetonitrile and acetic acid were purchased from Mallinckrodt Baker, Inc, (Phillipsburg, NJ, USA). Ethanol and ammonium hydroxide solutions ($NH_3 \cdot H_2O$) were purchased from Beijing Chemical Company (Beijing China); ultrapure water was obtained from a Milli-Q system (Millipore, Bedford, MA, USA). All other chemicals used were of ACS or HPLC grade.

2.2. Matrix Preparation.
A 20-mg/mL dilution of DAHC stock solution was prepared with ultrapure water. The DHAP matrix stock solution was prepared by dissolving 10 mg of DHAP in 1 mL of anhydrous ethanol. A series of modified DHAP/DAHC matrix solutions were prepared by mixing DHAP stock solution with DAHC stock solution at different ratios from 10 : 1 to 1 : 80 (v/v). The DHB/phosphoric acid (PA) matrix was prepared as previously reported [19]; we modified this previous protocol by dissolving 20 mg of DHB in 1 mL of 50% aqueous acetonitrile containing 1% phosphoric acid.

2.3. Sample Preparation.
α-Casein and β-casein (100 nM in 50 mM ammonium bicarbonate buffer at pH 8.3) were digested overnight at 37°C with trypsin at an enzyme/substrate ratio of 1 : 25 (w/w). After digestion, TFA was added into the sample solution at 1.0% to quench the reaction. The digest was diluted with 60% acetonitrile containing 0.1% TFA to produce a series of sample solutions at concentrations of 1 pmol/μL, 200 fmol/μL, 50 fmol/μL, 20 fmol/μL, and 10 fmol/μL. A stock solution containing standard peptides **p**T (1 mM) and **p**Y(1 mM) was prepared using ultra pure water. The solution was diluted with 60% acetonitrile containing 0.1% TFA to 1 pmol/μL, 200 fmol/μL, 50 fmol/μL, 20 fmol/μL, and 10 fmol/μL. All of the peptide solutions prepared for DHB/PA measurements were diluted with 1% phosphoric acid.

2.4. Human Histone H1 Phosphorylation and Digestion.
The phosphorylation reaction between histone H1 and CDK1 was carried out in vitro. Briefly, the reaction solution was prepared with 3 μL 10 × CDK1 reaction buffer, 0.15 μL CDK1 (150 U), 4 μL histone H1 (1 μg/μL), 100 μM ATP, and 13 μL water, then incubated at 30°C for 45 minutes. The phosphorylated histone H1 was then separated using SDS-PAGE (12%) and stained with Coomassie Brilliant Blue using microwave-assisted methods [28]. The in-gel digestion of phosphorylated histone H1 was performed as follows: the band of histone H1 was excised from the polyacrylamide gel, washed twice with water, and destained with 40% acetonitrile/50 mM NH_4HCO_3. The gel pieces were dehydrated with 100% acetonitrile and dried for 5 min using a speedvac. Disulfide bonds were reduced with DTT (10 mM, 56°C, 45 minutes), and the free sulfhydryl groups were alkylated with iodoacetamide (55 mM, 25°C, 60 minutes in the dark). Gel pieces were washed with 50 mM NH_4HCO_3, 50% acetonitrile/50 mM NH_4HCO_3, and dehydrated with 100% acetonitrile. After drying with a speedvac, the gel was rehydrated using 100 ng/μL trypsin (50 mM NH_4HCO_3, pH 8.3) on ice for 30 minutes, and the digestion was carried out at 37°C for 60 minutes and then quenched with 1.0% TFA. The tryptic peptides were extracted twice with 60% acetonitrile containing 0.1% TFA, and then the combined digest solution was concentrated to 10 μL under vacuum.

2.5. Sample Analysis by MALDI-TOF-MS and MALDI-TOF2-MS.
MALDI-TOF-MS analysis of peptides was performed using an AXIMA-CFR *plus* MALDI-TOF mass spectrometer (Shimadzu/Kratos, Manchester, UK) equipped with a pulsed nitrogen laser operated at 337 nm. Positive/negative ion MALDI mass spectra were acquired in the reflectron mode under the following parameters: ion source, 20/−20 kv, lens, 6.3/−6.3 kv, pulsed extraction, −2.5/2.5 kv, reflection, 25/−25 kV. High-energy CID MS/MS was performed with the AXIMA-TOF[2] MALDI mass spectrometer (Shimadzu/Kratos, Manchester, UK), equipped with the same pulsed 337-nm nitrogen laser. Operation parameters were: ion source, 20.0 kV; lens, 6.5 kV; pulsed extraction, −2.5 kV; and reflectron, 24.4 kV.

To prepare for MALDI-TOF-MS analysis, 1 μL of peptide solution was mixed with 3 μL of matrix solution in an Eppendorf tube, and 1 μL of the peptide/matrix solution was spotted onto the MALDI sample plate and then crystallized either under vacuum (when DHAP/DAHC was used as matrix) or in the air (when DHB/PA was used as matrix).
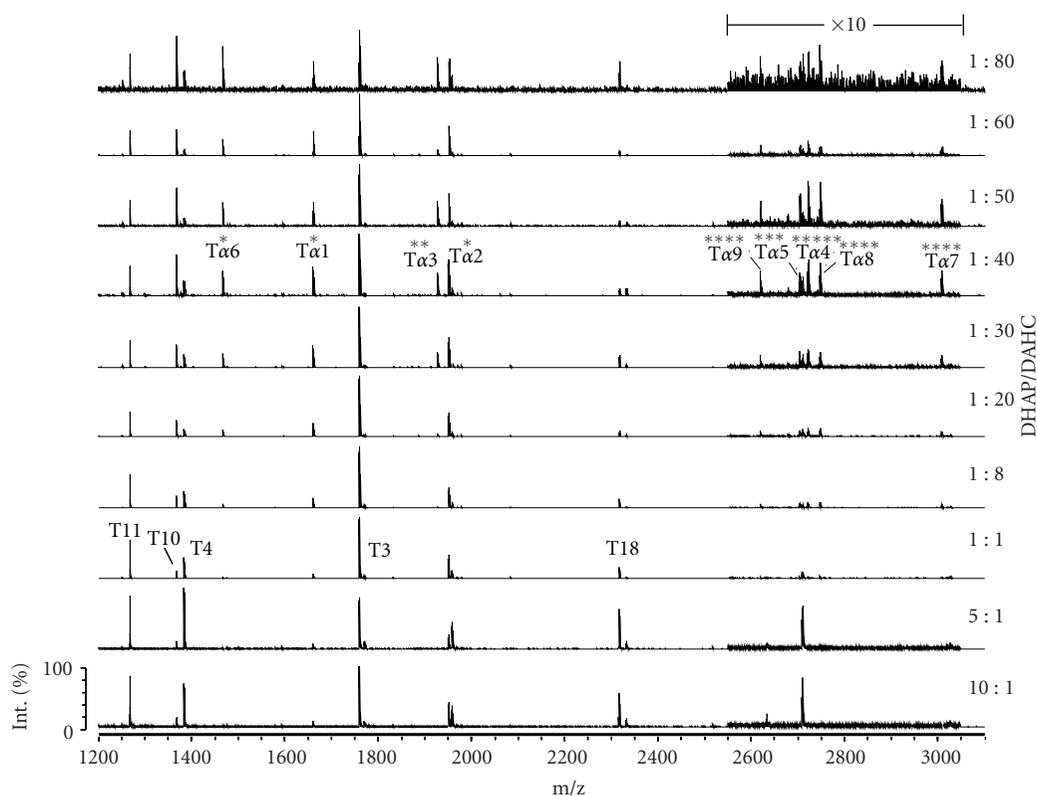
Figure 1: The MALDI mass spectra of the tryptic digest of α-casein (250 fmol on target) measured in positive ion mode with DHAP/DAHC matrix at different ratios (10 : 1 to 1 : 80). The signal intensity is magnified 10-fold in the m/z range from 2575 to 3025. The number of sites on each phosphopeptide is equal to the number of asterisks (*) shown.

Measurements using DHAP/DAHC were performed by detecting 15–20 different positions on the broad outer zone (see discussion below) for each sample to accumulate a total of 200 profiles (5 shots per profile). For measurements with DHB/PA, a total of 200 profiles (5 shots per profile) were accumulated by searching for 10–15 hot spots for each sample. All experiments were repeated three times.

The standard peptide **p**T (250 fmol on target) was used to determine sample homogeneity of the matrix. On each sample spot, 196 (14 × 14) positions were selected manually to ensure a full overview over the sample preparation. For each position, 50 profiles (5 shots per profile) were accumulated in positive ion mode. The signal intensities of **p**T in each position were recorded for further analysis.

Mass spectrometry data were processed with Launchpad 2.7.1 software (Shimadzu/Kratos, Manchester, UK). The data of the CID MS/MS spectra of phosphopeptides from histone H1 were compared to the SwissProt 56.6 protein sequence database (*Homo sapiens*) using the on-line MAS-COT database search engine (Matrix Science, London, UK) with the following parameters: trypsin with three missed cleavages; average mass values; peptide mass tolerance: ±0.8 Da; fragment mass tolerance: ±1.0 Da; fixed modifications: Carbamidomethyl (C), variable modifications: Oxidation

(M), phosphor (ST). Only the peptides with ions score above 38 were accepted as significant matches.

3. Results and Discussion

3.1. Optimization of DHAP/DAHC Matrix Combination. The solution composition of the matrix is a factor that significantly influences the ionization efficiency and the quality of MALDI mass spectra [29]. Therefore, we prepared a series of different ratios of DHAP (10 mg/mL) and DAHC (20 mg/mL) matrix solutions (10 : 1, 5 : 1, 1 : 1, 1 : 4, 1 : 8, 1 : 16, 1 : 20, 1 : 30, 1 : 40, 1 : 50, 1 : 60, and 1 : 80 v/v) and investigated these matrix combinations using the tryptic digest of α-casein, which contains nine phosphopeptides (Tα1-Tα9) and some nonphosphopeptides (such as T4, T10, T11, T18) (see Table 1). Figure 1 shows a comparison of the MALDI-TOF-MS analysis of the tryptic digest of α-casein (250 fmol on target) using matrix DHAP/DAHC at different ratios in positive ion mode. The signal intensity of phospho-peptides (especially for multiple phosphorylated peptides) changed dramatically with the ratio of DHAP/DAHC. It was clear that all phosphopeptides could be easily detected with a low DHAP/DAHC matrix ratio in the range of 1 : 30 to 1 : 60. In an optimal matrixsolution, the concentration

(a)

(b)

| ■ 300-400 | □ 100-200 |
| ■ 200-300 | □ 0-100 |

(c)

| ■ 300-400 | □ 100-200 |
| ■ 200-300 | □ 0-100 |

(d)

(e)
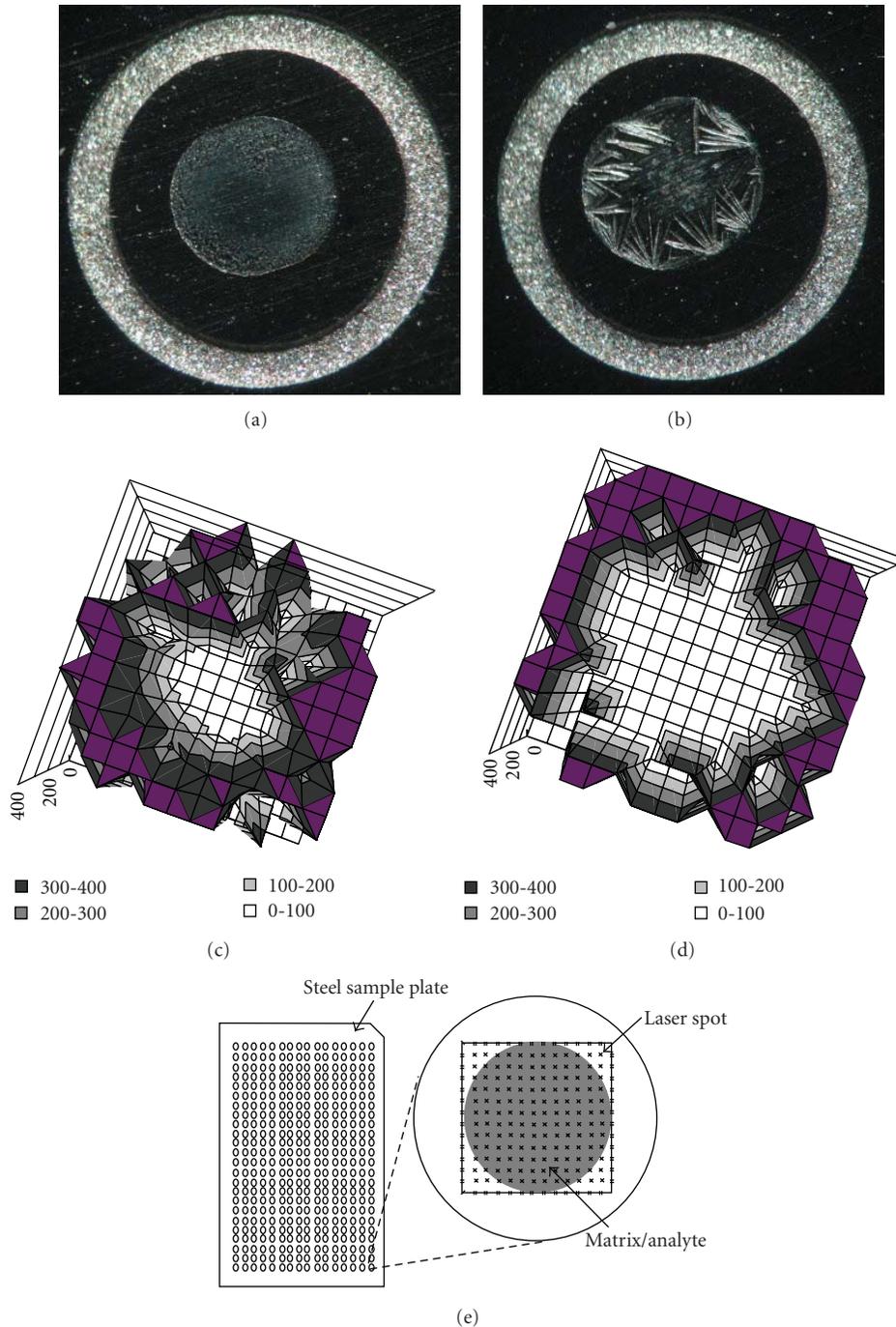
FIGURE 2: Top two views show the MALDI samples containing the phosphopeptide **p**T using (a) DHAP/DAHC matrix and (b) DHB/PA matrix. The two center pictures show the signal intensities of the phosphopeptides **p**T (250 fmol on target) measured with (c) DHAP/DAHC matrix and (d) DHB/PA matrix. The images only show a cut at signal intensity of 400 mv for clarity. The positions with signal intensities above 400 mv are indicated in purple. The bottom row (e) shows a schematic of the data acquisition method for investigating sample homogeneity of the matrix. In total, 196 positions were selected as a 14×14 spots array to cover the crystalline matrix/analyte layer, and 50 profiles (5 shots per profiles) were accumulated at each position.

of DHAP was 1.6–3.3 mg/mL, which is lower than the concentration of 10–15 mg/mL used by other groups [24, 26, 27]. Once the DHAP/DAHC ratio was down to 1 : 80, most multiple phosphorylated peptides could no longer be detected, likely because there were not sufficient numbers of

DHAP molecules to transfer the laser energy to the peptides to be protonated.

Another interesting phenomenon apparent in Figure 1: the signal intensity of nonphosphopeptides was stable (T3), enhanced (T10) or weakened (T11, T4, T18) with the
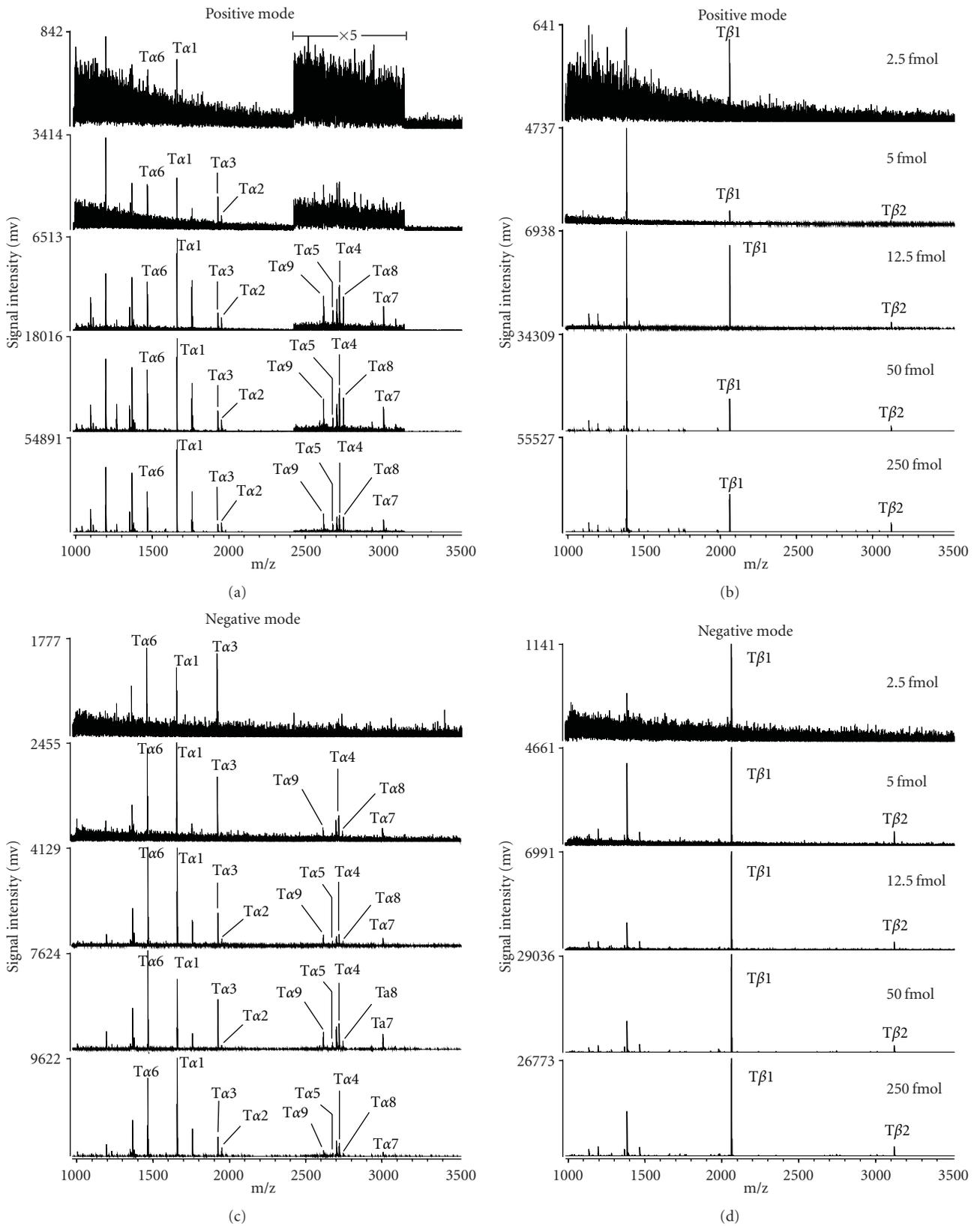
Figure 3: The MALDI mass spectra of the tryptic digest of (a) and (c) α-casein and (b) and (d) β-casein measured in positive (upper) and negative (lower) ion modes with the DHAP/DAHC matrix. The amount of the tryptic peptides mixture ranged from 5 fmol to 250 fmol on target. The signal intensity is magnified 5-fold in the m/z range from 2400 to 3150 in section a.
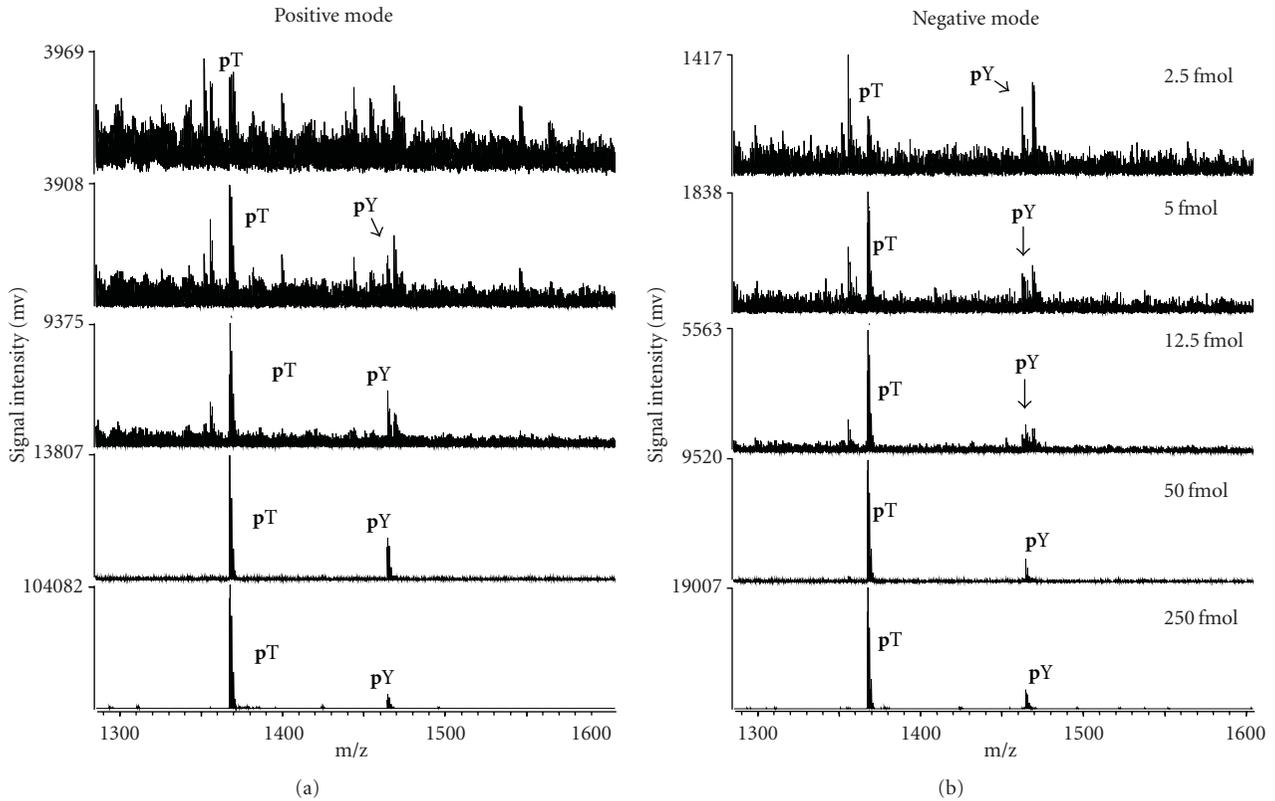
Figure 4: The MALDI mass spectra of two synthesized phosphopeptides **p**T and **p**Y mixed at equimolarity from 5 to 250 fmol on target, which were measured in positive (a) and negative (b) ion mode with DHAP/DAHC, respectively.

Table 1: The peptides investigated in this study*.

Peptide ID	Sequence No.	Sequence	$[M + H]^+_{\text{Theo}}$	B+B
Tα1	aa 106–119	VPQLEIVPN**pS**AEER	1660.79	260
Tα2	aa 104–119	KYKVPQLEIVPN**pS**AEER	1951.95	−710
Tα3	aa 43–58	DIG**pS**E**pS**TEDQAMEDIK	1927.68	3960
Tα4	aa 59–79	QMEAE**pS**I**pSpSpS**EEIVPN**pS**VEQK	2720.90	3680
Tα5	aa 37–58	VNEL**pS**KDIG**pS**E**pS**TEDQAMEDIK	2678.02	3840
Tα6	aa 138–149	TVDME**pS**TEVFTK	1466.60	−210
Tα7	aa 46–70	NANEEEYSIG**pSpSpS**EE**pS**AEVATEEVK	3008.02	7100
Tα8	aa 1–21	KNTMEHV**pSpSpS**EESII**pS**QETYK	2747.00	2810
Tα9	aa 2–21	NTMEHV**pSpSpS**EESII**pS**QETYK	2618.91	2350
Tβ1	aa 33–48	FQ**pS**EEQQQTEDELQDK	2061.82	6200
Tβ2	aa 1–25	RELEELNVPGEIVE**pS**L**pSpSpS**EESITR	3122.26	−220
T10	aa 81–91	ALNEINQFYQK	1367.70	−750
T4	aa 23–34	FFVAPFPEVFGK	1384.73	−5530
T3	aa 8–22	HQGLPQEVLNENLLR	1759.59	−590
T11	aa 91–100	YLGYLEQLLR	1267.71	−6480
T18	aa 132–150	EPMIGVNQELAYFYPELFR	2316.14	−6900
pT	—	VNQIG**pT**LSESIK	1368.44	−310
pY	—	SGSLHRI**pY**THQS	1465.48	2300

*Phosphopeptides in tryptic digest of βcasein and αcasein are denoted as Tβ1~2 and Tα1~9, respectively. The sites of phosphorylation are indicated with the letter "p" in the listed peptide sequences. T3, T4, T11, and T18 are nonphosphorylated peptide from αcasein S1. T10 is a nonphosphorylated peptide from α-casein S2. **p**T and **p**Y are the synthesized standard phosphopeptides. $[M+H]^+$ denotes the theoretical monoisotopic m/z mass. "B+B" means Bull and Breese Index for peptides calculated by the Masslynx software (Waters/Micromass).
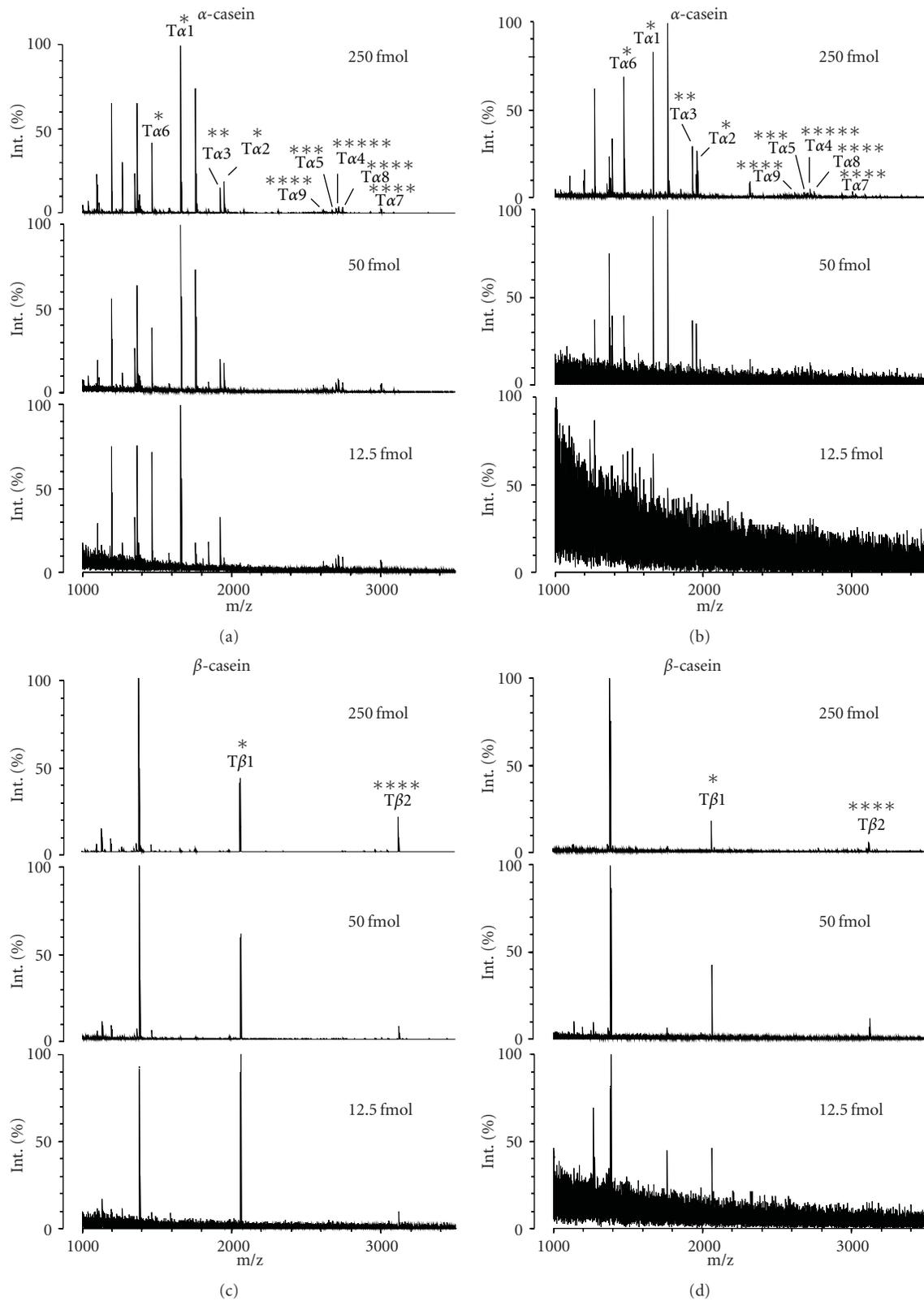
FIGURE 5: Tryptic digests of α- (upper) and β-casein (lower) (12.5–250 fmol on target) were detected with DHAP/DAHC (a) and (c) and DHB/PA (b) and (d) by MALDI-TOF MS in positive mode. DHAP/DAHC positions were selected randomly in the outer ring of the sample; for DHB/PA, only signals in hot spots were collected. In total, 200 profiles were accumulated for each sample. The number of sites on each phosphopeptide was equal to the number of asterisks (*).

(a)

(b)

(c)
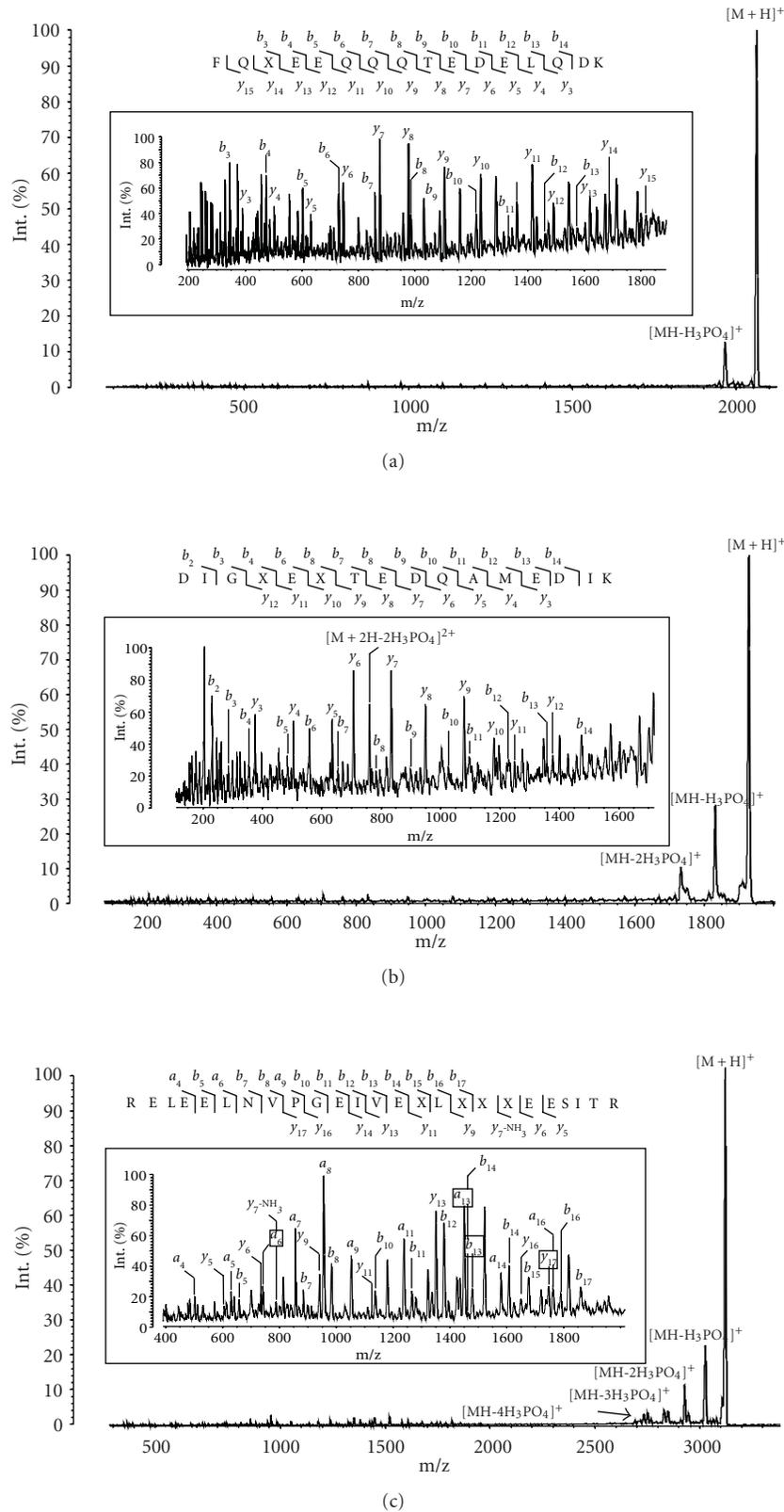
Figure 6: The MALDI MS/MS spectra of phosphopeptides (a) Tβ1, (b) Tα3, and (c) Tβ2 (250 fmol for each on target) from α-casein and β-casein were obtained by MALDI-TOF/TOF MS. The sequential losses of H_3PO_4 from the parent ions are noted. The fragment patterns of the peptides are indicated in the magnified MS/MS spectra. "**X**" in the amino sequence shows a dehydroalanine residue converted from a phosphoserine residue by beta-elimination of H_3PO_4.
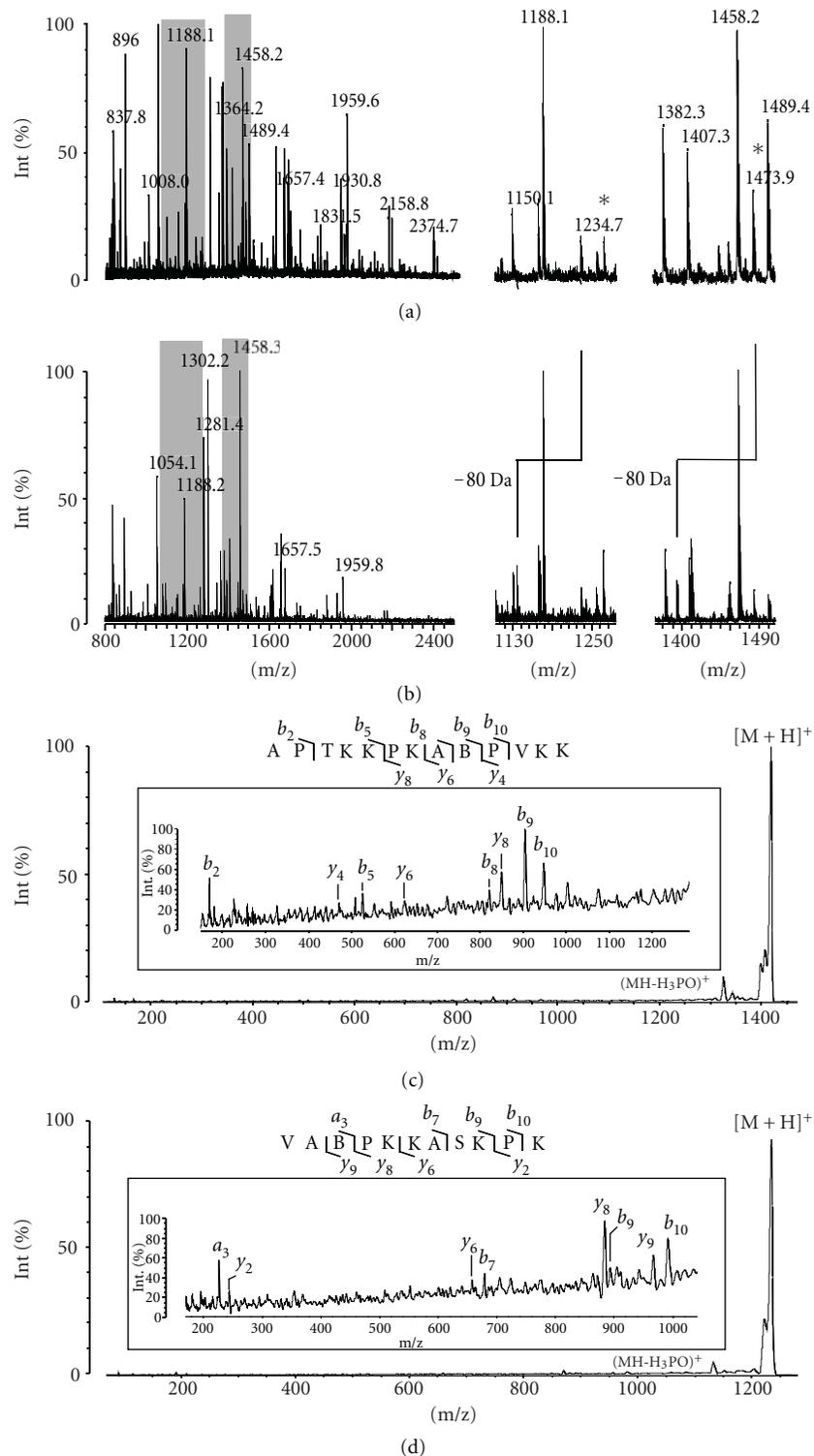
Figure 7: The MALDI MS analysis of phosphopeptides from CDK1-treated human histone H1. MS spectra of the tryptic peptides from CDK1-treated histone H1 with the untreated (Figure 7(a)) and alkaline phosphatase-treated histone H1 (Figure 7(b)). The two panels on the right show the magnified spectra to indicate the two phosphopeptides labeled with asterisks (*). The MALDI-TOF/TOF-MS analysis of phosphorylation sites on the two monophosphopeptides VApTPKKASKPK, m/z 1234.7 (c), APTKKPKApTPVKK, m/z 1473.9 (d) from the CDK1-treated human histone H1. The neutral-loss peak of phosphopeptide was noted as $[MH-H_3PO_4]^+$. The fragment patterns of peptides were shown in the magnified MS/MS spectra. "B" in the amino sequence indicates a dehydroamino-2-butyric acid residue converted from a phosphothreonine residue by beta-elimination of H_3PO_4. All the spectra were detected using DHAP/DAHC matrix in the positive mode.

change of the DHAP/DAHC ratio. We calculated the Bull and Breese Index (BB Index) [19] of these peptides, an index generally used to reflect the hydrophobic/hydrophilic nature of peptides. Higher BB Index values are associated with more hydrophilic peptides. For the phosphopeptide, we calculated the BB Index by replacing the phosphoserine or phosphothreonine residues with glutamic acid [19]. By comparing the BB Index of peptides listed in Table 1, we found that the phosphopeptides were nearly hydrophilic peptides with high BB Indices and that the nonphosphopeptides for which signal intensity was weakened at low DHAP/DAHC ratio were highly hydrophobic with lower BB Index (T4 : −5530, T11 : −6480, T18 : −6900). The nonphosphopeptide T10 for which the ion signal was enhanced at lower ratios of DHAP/DAHC was found to be slightly hydrophobic with a BB Index of −750. These findings indicated that the optimal DAHC/DHAP matrix tended to detect more hydrophilic peptides. Most phosphopeptides are hydrophilic, so we propose that the low ratio of DHAP/DAHC matrix could increase cocrystallization between matrix and phosphopeptides and further improve ionization efficiency of phosphopeptide in MALDI MS.

3.2. Sample Homogeneity.

In this study, the phosphopeptide **p**T (250 fmol on target) was used to investigate the sample homogeneity of optimized DHAP/DAHC matrix, and for a comparison, we chose matrix combination of 2,5-DHB with 1% phosphoric acid, a currently popular matrix used for detection of phosphopeptide [30–32]. It was apparent from the photographs of the MALDI sample spots that there was a dramatic difference between the optimized DHAP/DAHC (Figure 2(a)) and DHB/PA (Figure 2(b)). Small crystals formed in a DHAP/DAHC sample spot while larger, needle-like crystals formed in a DHB/PA sample spot. Furthermore, 196 (14 × 14) positions on the target were measured covering the crystalline layer for a full overview (Figure 2(e)), and the MS signal intensities of pT were monitored. The majority of the outer zones of the crystalline samples exhibited high signal intensities, except for a small center zone when DHAP/DAHC used as matrix (Figure 2(c)). In comparison, the crystalline DHB/PA matrix exhibited the hot-spot phenomenon, signals were primarily observed in the narrow, inhomogeneous ring at the edge of the sample layer, and the inner position almost did not exhibit any signal (Figure 2(d)). Therefore, the samples prepared with DHAP/DAHC were more homogeneous than those prepared with a DHB/PA matrix, reducing the need for time-consuming hot-spot searches. Based on the sample homogeneity of DHAP/DAHC and DHB/PA, in the next experiments, the data acquisition with DHAP/DAHC was carried out by randomly measuring the outer zone of the sample layer, while data were acquired from the DHB/PA matrix by searching for hot spots of sample.

3.3. Sensitivity Measurement of Phosphopeptides with the DHAP/DAHC Matrix.

In order to further evaluate the measurement sensitivity of phosphopeptides with the optimal

DHAP/DAHC matrix, the tryptic digest of α-casein and β-casein with 2.5–250 fmol on the target were detected by MALDI-TOF MS in both positive and negative ion mode. Figure 3 shows that all eleven phosphopeptides from α-casein and β-casein could be detected with signal-to-noise (S/N) ratios greater than 3 at the low amount of 12.5 fmol in the positive mode (a) and (b) and detected at a lower amount of 5 fmol in the negative mode (c) and (d). The monophosphopeptides (Tα1 and Tβ1) could still be observed at 2.5 fmol on target in negative ion mode (Figure 3(c)). Phosphopeptides generally exhibited better S/N ratios in the MS spectra in negative ion mode than that in positive ion mode, although, consistent with previous reports [19], the absolute signal intensity of phosphopeptide was usually lower in negative ion mode. Because all tryptic phosphopeptides from α-casein and β-casein are serine-phosphorylated peptides, a threonine-phosphorylated peptide **p**T and a tyrosine-phosphorylated peptide **p**Y (2.5–250 fmol on target) were used to investigate whether the DHAP/DAHC matrix had a generic nature in detection of phosphopeptides. As expected, **p**T and **p**Y could be readily detected down to 5 fmol on the target in both positive and negative modes (Figure 4).

We further directly compared optimized DHAP/DAHC to DHB/PA. Tryptic digests of α-casein and β-casein (12.5–250 fmol on target) were detected using these two matrices (Figure 5). When using the DHAP/DAHC matrix, all the phosphopeptides could readily be detected at three different sample amounts. By comparison, when DHB/PA was used as the matrix, although all the phosphopeptides in 250 fmol digest (on target) could be detected with DHB/PA, the amount of digest went down to 50 fmol (on the target), and only Tα1, Tα2, Tα3, Tα6, Tβ1, and Tβ2 could be observed. Furthermore, when the amount of digest went down to 12.5 fmol, only Tα1 and Tβ1 were detected, with weak signals. Comparing these results, it was clear that compared with DHB/PA, the optimized DHAP/DAHC had comparative sensitivity in the detection of singly phosphorylated peptides (Tα1 and Tβ1), but greater sensitivity in the detection of multiply phosphorylated peptides, such as Tβ2 (4**p**) and Tα4 (5**p**). Herein, we suggest that sample should be analyzed immediately after sample preparation when optimized DHAP/DAHC matrix used, as the crystals might not last too long under atmospheric conditions and in vacuo.

3.4. Sequencing of Phosphopeptides with High Energy CID MS/MS.

We employed the AXIMA-TOF[2] MALDI mass spectrometer (Shimadzu/Kratos, Manchester, UK) to perform high energy CID MS/MS of phosphopeptides to characterize phosphorylation sites. Figure 6 shows the MS/MS spectra of three phosphopeptides Tβ1, Tα3, and Tβ2 from α-casein and β-casein (250 fmol on target) with DHAP/DAHC as matrix. The number of [MH-98]$^+$ or [MH-80]$^+$ peaks in the MS/MS spectra indicated the number of phosphate groups in each phosphorylated peptide, and the fragment ions including a, b, and y ions clearly show the phosphorylation sites of the phosphopeptides Tβ1, Tα3, and Tβ2. These results demonstrated the feasibility of MALDI-TOF/TOF

MS with the DHAP/DAHC matrix for the analysis of the phosphorylation sites of phosphopeptides, including singly and multiply phosphorylated peptides.

3.5. Application of the Optimized DHAP/DAHC Matrix in the Characterization of Phosphorylated Human Histone H1. Further investigating the practicality of the DHAP/DAHC matrix, we used this approach to characterize human histone H1 phosphorylated with CDK1. Experimentally, phosphorylated histone H1 was separated by SDS-PAGE and in-gel digested by trypsin. In order to generate appropriate peptide sizes for MALDI-MS analysis, since the sequence of histone proteins is known to contain many lysine and arginine residues, the trypsin digestion time was optimized and reduced to 1 hour. Figure 7(a) shows the results of the MALDI-TOF-MS analysis of the tryptic digest of the phosphorylated histone H1 without the desalting step, directly detected with the optimized DHAP/DAHC matrix in positive ion mode. After performing data processing using the on-line MASCOT search engine, 31 peptides, including 2 possible phosphopeptides (in Figure 7(a) labeled with "*"), were assigned to human histone H1, and 81% of the sequence was covered. These two peptides corresponded to the two monophosphopeptides of histone H1, comprising amino acids 117–127 (VATPKKASKPK, 1234.7 m/z; possible phosphorylation sites indicated with underlined letters) and 133–145 (APTKKPKATPVKK, 1473.9 m/z). The tryptic sample of H1 was also analyzed with DHB/PA, but the two peaks of the phosphopeptide were lower than those in Figure 7(a) (see Figure S1). To validate the MASCOT searching result for the two phosphorylated peptides, we compared the mass spectra of tryptic peptides of H1 before and after treatment with alkaline phosphatase, which can cleave the phosphate group from phosphopeptides. It was clear that the two peptide peaks (marked with asterisks in Figure 7(a)) had disappeared and two new peaks with a mass shift of 80 Da (HPO_3 = 80 Da) were visible (Figure 7(b)). This demonstrated that the two peptides (marked with asterisks in Figure 7(a) were singly phosphorylated peptides. To identify the phosphorylation sites, we used MALDI-TOF/TOF MS to perform MS/MS analysis of the two peptides (Figure 7(c) and 7(d)). Only one neutral-loss peak corresponding to $[MH-H_3PO_4]^+$ indicated that both phosphopeptides were monophosphopeptides. After a MASCOT search, the phosphorylation sites of the two phosphopeptides were determined to be APTKKPKApTPVKK (pT indicating the phosphothreonine) and VApTPKKASKPK with ion scores of 53 and 50, respectively (scores exceeding 38 were accepted as significant matches).

In general, the protein phosphorylation sites by a particular protein kinase shared a set of "consensus sequence", which is necessary and sufficient for recognition by the kinase [33]. To further validate the phosphorylation sites of histone H1 identified with MALDI-TOF/TOF MS, we compared the two phosphorylated peptide sequences with the consensus sequence pS/pT-P-X-R/K, most frequently recognized by CDK1 [34, 35], and found that both phosphopeptide sequences were consistent with the consensus sequence.

These results thus demonstrate that the DHAP/DAHC matrix was robust and effective in analyzing the protein phosphorylation of the biological sample by MALDI-MS.

4. Conclusion

By optimizing the matrix solution composition, we found that a low ratio of DHAP/DAHC formulation was more efficient in detecting phosphopeptides than earlier protocols. Compared with DHB/PA, the optimized DHAP/DAHC exhibited higher sample homogeneity, and the phosphopeptides in tryptic digests of α- and β-casein without desalting and phosphopeptide enrichment could be measured with a higher sensitivity. Further, the optimized DHAP/DAHC showed high energy CID MS/MS analysis of the phosphorylation sites of phosphopeptides, including multiple phosphorylated peptides, and we successfully applied this method to the characterization of the phosphorylation sites of CDK1-treated human histone H1.

References

[1] J. Posada and J. A. Cooper, "Molecular signal integration. Interplay between serine, threonine, and tyrosine phosphorylation," *Molecular Biology of the Cell*, vol. 3, no. 6, pp. 583–592, 1992.

[2] T. Hunter, "Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling," *Cell*, vol. 80, no. 2, pp. 225–236, 1995.

[3] T. Hunter, "Signaling—2000 and beyond," *Cell*, vol. 100, no. 1, pp. 113–127, 2000.

[4] P. Cohen, "The regulation of protein function by multisite phosphorylation—a 25 year update," *Trends in Biochemical Sciences*, vol. 25, no. 12, pp. 596–601, 2000.

[5] M. Mukherji, "Phosphoproteomics in analyzing signaling pathways," *Expert Review of Proteomics*, vol. 2, no. 1, pp. 117–128, 2005.

[6] T.-T. Yip and T. W. Hutchens, "Mapping and sequence-specific identification of phosphopeptides in unfractionated protein digest mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry," *FEBS Letters*, vol. 308, no. 2, pp. 149–153, 1992.

[7] A. G. Craig, C. A. Hoeger, C. L. Miller, T. Goedken, J. E. Rivier, and W. H. Fischer, "Monitoring protein kinase and phosphatase reactions with matrix-assisted laser desorption/ionization mass spectrometry and capillary zone electrophoresis: comparison of the detection efficiency of peptide-phosphopeptide mixtures," *Biological Mass Spectrometry*, vol. 23, no. 8, pp. 519–528, 1994.

 [8] P.-C. Liao, J. Leykam, P. C. Andrews, D. A. Gage, and J. Allison, "An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry," *Analytical Biochemistry*, vol. 219, no. 1, pp. 9–20, 1994.

 [9] R. S. Annan and S. A. Carr, "Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry," *Analytical Chemistry*, vol. 68, no. 19, pp. 3413–3421, 1996.

[10] S. A. Carr, M. J. Huddleston, and R. S. Annan, "Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry," *Analytical Biochemistry*, vol. 239, no. 2, pp. 180–192, 1996.

[11] M. Busman, K. L. Schey, J. E. Oatis Jr., and D. R. Knapp, "Identification of phosphorylation sites in phosphopeptides by positive and negative mode electrospray ionization-tandem mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 7, no. 3, pp. 243–249, 1996.

[12] R. S. Annan and S. A. Carr, "The essential role of mass spectrometry in characterizing protein structure: mapping posttranslational modifications," *Protein Journal*, vol. 16, no. 5, pp. 391–402, 1997.

[13] K. Stühler and H. E. Meyer, "MALDI: more than peptide mass fingerprints," *Current Opinion in Molecular Therapeutics*, vol. 6, no. 3, pp. 239–248, 2004.

[14] M. Khan, H. Takasaki, and S. Komatsu, "Comprehensive phosphoproteome analysis in rice and identification of phosphoproteins responsive to different hormones/stresses," *Journal of Proteome Research*, vol. 4, no. 5, pp. 1592–1599, 2005.

[15] G. Yan, L. Li, Y. Tao, et al., "Identification of novel phosphoproteins in signaling pathways triggered by latent membrane protein 1 using functional proteomics technology," *Proteomics*, vol. 6, no. 6, pp. 1810–1821, 2006.

[16] H. Steen, J. A. Jebanathirajah, J. Rush, N. Morrice, and M. W. Kirschner, "Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements," *Molecular and Cellular Proteomics*, vol. 5, no. 1, pp. 172–181, 2006.

[17] P. Cohen, "The role of protein phosphorylation in human health and disease," *European Journal of Biochemistry*, vol. 268, no. 19, pp. 5001–5010, 2001.

[18] D. Arnott, M. A. Gawinowicz, R. A. Grant, et al., "ABRF-PRG03: phosphorylation site determination," *Journal of Biomolecular Techniques*, vol. 14, no. 3, pp. 205–215, 2003.

[19] S. Kjellström and O. N. Jensen, "Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins," *Analytical Chemistry*, vol. 76, no. 17, pp. 5109–5117, 2004.

[20] A. Tholey, "Ionic liquid matrices with phosphoric acid as matrix additive for the facilitated analysis of phosphopeptides by matrix-assisted laser desorption/ionization mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 20, no. 11, pp. 1761–1768, 2006.

[21] H. Zhou, S. Xu, M. Ye, et al., "Zirconium phosphonate-modified porous silicon for highly specific capture of phosphopeptides and MALDI-TOF MS analysis," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2431–2437, 2006.

[22] X. Yang, H. Wu, T. Kobayashi, R. J. Solaro, and R. B. Van Breemen, "Enhanced ionization of phosphorylated peptides during MALDI TOF mass spectrometry," *Analytical Chemistry*, vol. 76, no. 5, pp. 1532–1536, 2004.

[23] J. M. Asara and J. Allison, "Enhanced detection of phosphopeptides in matrix-assisted laser desorption/ionization mass spectrometry using ammonium salts," *Journal of the American Society for Mass Spectrometry*, vol. 10, no. 1, pp. 35–44, 1999.

[24] J. J. Gorman, B. L. Ferguson, and T. B. Nguyen, "Use of 2,6-dihydroxyacetophenone for analysis of fragile peptides, disulphide bonding and small proteins by matrix-assisted laser desorption/ionization," *Rapid Communications in Mass Spectrometry*, vol. 10, no. 5, pp. 529–536, 1996.

[25] T. Nabetani, K. Miyazaki, Y. Tabuse, and A. Tsugita, "Analysis of acidic peptides with a matrix-assisted laser desorption/ionization mass spectrometry using positive and negative ion modes with additive monoammonium phosphate," *Proteomics*, vol. 6, no. 16, pp. 4456–4465, 2006.

[26] C.-F. Xu, Y. Lu, J. Ma, M. Mohammadi, and T. A. Neubert, "Identification of phosphopeptides by MALDI Q-TOF MS in positive and negative ion modes after methyl esterification," *Molecular and Cellular Proteomics*, vol. 4, no. 6, pp. 809–818, 2005.

[27] C.-F. Xu, H. Wang, D. Li, X.-P. Kong, and T. A. Neubert, "Selective enrichment and fractionation of phosphopeptides from peptide mixtures by isoelectric focusing after methyl esterification," *Analytical Chemistry*, vol. 79, no. 5, pp. 2007–2014, 2007.

[28] V. J. Nesatyy, A. Dacanay, J. F. Kelly, and N. W. Ross, "Microwave-assisted protein staining: mass spectrometry compatible methods for rapid protein visualisation," *Rapid Communications in Mass Spectrometry*, vol. 16, no. 4, pp. 272–280, 2002.

[29] S. L. Cohen and B. T. Chait, "Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins," *Analytical Chemistry*, vol. 68, no. 1, pp. 31–37, 1996.

[30] T. E. Thingholm, T. J. D. Jørgensen, O. L. Jensen, and M. R. Larsen, "Highly selective enrichment of phosphorylated peptides using titanium dioxide," *Nature Protocols*, vol. 1, no. 4, pp. 1929–1935, 2006.

[31] G. Luo, A. Gruhler, Y. Liu, O. N. Jensen, and R. C. Dickson, "The sphingolipid long-chain base-Pkh1/2-Ypk1/2 signaling pathway regulates eisosome assembly and turnover," *Journal of Biological Chemistry*, vol. 283, no. 16, pp. 10433–10444, 2008.

[32] C. Krintel, P. Osmark, M. R. Larsen, S. Resjö, D. T. Logan, and C. Holm, "Ser649 and Ser650 are the major determinants of protein kinase A-mediated activation of human hormone-sensitive lipase against lipid substrates," *PLoS ONE*, vol. 3, no. 11, article e3756, 2008.

[33] P. J. Kennelly and E. G. Krebs, "Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases," *Journal of Biological Chemistry*, vol. 266, no. 24, pp. 15555–15558, 1991.

[34] L. M. Stevenson-Lindert, P. Fowler, and J. Lew, "Substrate specificity of CDK2-cyclin A: what is optimal?" *Journal of Biological Chemistry*, vol. 278, no. 51, pp. 50956–50960, 2003.

[35] Z. Songyang, S. Blechner, N. Hoagland, M. F. Hoekstra, H. Piwnica-Worms, and L. C. Cantley, "Use of an oriented peptide library to determine the optimal substrates of protein kinases," *Current Biology*, vol. 4, no. 11, pp. 973–982, 1994.

Research Article

Two-Dimensional Liquid Chromatography Technique Coupled with Mass Spectrometry Analysis to Compare the Proteomic Response to Cadmium Stress in Plants

Giovanna Visioli, Marta Marmiroli, and Nelson Marmiroli

Division of Genetics and Environmental Biotechnologies, Department of Environmental Sciences, University of Parma, Viale G.P. Usberti 11/A, 43100 Parma, Italy

Correspondence should be addressed to Nelson Marmiroli, nelson.marmiroli@unipr.it

Plants are useful in studies of metal toxicity, because their physiological responses to different metals are correlated with the metal exposure dose and chemical state. Moreover a network of proteins and biochemical cascades that may lead to a controlled homeostasis of metals has been identified in many plant species. This paper focuses on the global protein variations that occur in a *Populus nigra* spp. clone (Poli) that has an exceptional tolerance to the presence of cadmium. Protein separation was based on a two-dimensional liquid chromatography technique. A subset of 20 out of 126 peaks were identified as being regulated differently under cadmium stress and were fingerprinted by MALDI-TOF. Proteins that were more abundant in the treated samples were located in the chloroplast and in the mitochondrion, suggesting the importance of these organelles in the response and adaptation to metal stress.

1. Introduction

Cadmium (Cd) is a widespread element in the environment, mainly from anthropogenic sources. At low concentrations, it is toxic to microorganisms, plants, and animals. In humans, Cd intake is normally by ingestion or inhalation, with the majority of ingested Cd coming from contaminated foods (meat and plants) and water. Thus the transfer of Cd from the soil into plants and from Cd deposited from the troposphere onto edible plants accounts for the majority of human Cd intake [1].

In agriculture, Cd pollution is an increasing problem due to soil amendments and to the intensive use of phosphate fertilizers. In natural environments, Cd is present at concentrations of 0.1–0.5 mg kg^{-1}, but it may reach concentrations as high as 150 mg kg^{-1} at contaminated sites. The high solubility of Cd causes rapid distribution in the environment, where it is immediate availability to plants and from these primary producers can move up through the entire food chain [2]. On the other hand, the ability of some plants to absorb and accumulate significant quantities of metals in roots, shoots, and leaves offers a unique opportunity to remove inorganic pollutants from contaminated media [3].

Laboratory and field trials have found that many clones from the genus *Populus* seem promising for the removal of Cd from soil [3, 4]. Poplar plants are ecologically and economically attractive due to their fast growth, high biomass production, and easy propagation through vegetative cuttings. Moreover poplars are used for different industrial applications, including renewable energy production.

The use of plants to remediate metal pollution [5] relies on an understanding of the molecular mechanisms involved in the uptake, translocation, and sequestration of metals by plants. Thus, research focusing on genomic and transcriptomic analyses in plants has been conducted after metal exposure in field trials [6].

A better knowledge of the plant proteome, as a way of understanding phenotypic plasticity and adaptability in plants, is required to effectively exploit plant biological resources [7]. Up to a few years ago, the study of proteomic

variability was restricted to a few model plants, such as *Arabidopsis thaliana*, under a wide range of stresses. Recently, comparative analysis of proteome profiles in plants following metal exposure has attracted greater interest [8, 9].

Comparative proteomics rests on two main approaches: (i) the first is a "gel-based" method, which uses a two-dimensional electrophoresis separation of proteins (2D-PAGE) [10, 11] and (ii) the second is a "gel-less" method, in which proteins are separated by two-dimensional liquid chromatography (2D-LC), a technique which promises to extend the range of protein separation [12, 13]. Using comparative proteomics, protein variations of the *Populus nigra* clone "Poli" grown without Cd were compared to clones grown with a 50 μM CdSO$_4$ treatment, which was sublethal for this clone [14]. The protein patterns obtained from plant samples of the two experimental conditions were compared using a semiquantitative 2D-LC technique. Three sets of proteins were characterized: (i) more abundant in the treated sample in respect to the control, (ii) less abundant in the treated sample in respect to the control, and (iii) equally abundant in both treated and control samples [13]. Proteins with altered expression ((i) and (ii)) were digested with trypsin and identified by Matrix-assisted LASER desorption / ionization time of flight mass spectrometry (MALDI-TOF/MS). The combination of a software that could quantify protein differences after 2D-LC and identify proteins, with MS technique, gave both qualitative and semiquantitative evidence of some of the proteome changes due to Cd treatment in this particular poplar clone.

2. Materials and Methods

2.1. Plant Material. Stem cuttings of *Populus nigra* clone "Poli" were grown hydroponically for three weeks in Hoagland's nutrient solution (one-third strength, pH 6.5 (mol l^{-1})) [15]. Cuttings were placed in a controlled climate chamber equipped with metal halide lamps (Powerstar HQI-TS; Osram, Munich, Germany), which provided a photon flux density of 300 μmol m^{-2} s^{-1} for 14 h at 25°C. During the 10 h dark period the temperature was 20°C. The relative humidity was 70%–80%. After three weeks, half of the plants were treated with 50 μM CdSO$_4$ for three more weeks. The hydroponic solution was controlled and maintained at a steady level with periodic supplements to compensate for the high transpiration rates of the plants. At the end of the experimental period, leaves from treated and untreated plants were sampled according to Zacchini et al. [14], frozen in liquid nitrogen, and stored at −80°C for protein extractions. This procedure was repeated three times using three cuttings from each treatment. Leaves were chosen for proteomic analysis because of their relative abundance in the total plant biomass and because their physiological behaviors (e.g., functions, photosynthesis, respiration, transpiration, etc), correlate strongly with the overall health of the whole plant. Detection of Cd uptake and translocation in poplars was previously described by Zacchini et al. [14].

2.2. D-LC Analysis. Five grams of frozen leaves from each treatment were finely ground with a mortar and pestle and then thawed in 4 ml of MgSO$_4$-based extraction buffer as described by Pirondini et al. [13]. Leaf samples from each treatment were extracted in triplicate for crude total proteins mixtures, as described below.

The crude mixture was sonicated for 10 min, and then the solution was centrifuged at 16 000× g for 5 min at 4°C. The pellet containing the larger cellular residues was discharged, and the supernatant was centrifuged at 16 000× g for 30 min at 4°C. The lighter phase contained the purified solubilised proteins. The purified proteins solution was loaded onto a desalted PD-10 column (GE-Healthcare, Uppsala, Sweden). Proteins were quantified determining the O.D. with a spectrophotometer (Beckman Coulter, Fullerton, CA) at 562 nm utilizing a BCA Protein Assay Kit (Novagen, Merck KGaA, Darmstadt, Germany). Protein separation was carried out using a 2D-LC technique (ProteomeLab PF 2D, Beckman Coulter, Fullerton, CA) following a procedure described by Pirondini et al. [13]. Equal amounts (1.3 mg) of total protein extract from leaves of each treatment (untreated or treated for 3 weeks with 50 μM CdSO$_4$) were loaded into the 1st dimension high-performance chromatofocusing (HPCF) column, where protein separation occurred by isoelectric point (pI). The resulting pH fractions were then injected into the 2nd dimension high-performance reversed-phase chromatography (HPRP) column, where separation occurred by protein hydrophobicity. The eluent from the 2nd dimension column was analyzed by U.V. detector for O.D. at 214 nm. This method provided a more consistent and sensitive detection of proteins via peptide bonds. Each protein was then represented by a Gaussian-shaped peak.

2.3. MALDI-TOF Analysis. Eluted proteins were manually collected from the column in single fractions every 0.2 min, starting from the beginning of the peak area. Half of each eluted peak was evaporated and suspended in an SDS-PAGE loading buffer and then loaded onto a 12% 1D-SDS-PAGE gel to check for the number of protein species (data not shown). Only peaks with a single band in the 1D-SDS-PAGE gel were evaporated by Speed Vac into a final volume of 10 μL for further analysis. Protein reduction was performed by incubating each fraction with 25 mM NH$_4$HCO$_3$ and 2 mM DTT in a water bath at 60°C for 1 h. The alkylation of the reduced sulfhydryl groups was carried out by adding 1 mM iodoacetamide (IAA), at 25°C, for 30 minutes in the dark. 1.5 μL of trypsin (125 μg mL^{-1}) in 50 mM NH$_4$HCO$_3$ was subsequently added to each fraction and the digestion was carried out at 37°C for 24 h. The digested samples were then purified with a ZipTipC18 (Millipore, Billerica, MA, USA) using the procedure recommended by the manufacturer. 1 μL of each purified peptide mixture was spotted directly onto a stainless steel MALDI target plate with 1 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid in 0.1% TFA : ACN (2 : 1, v/v). The solution was dried at room temperature and a spot was produced. Positively charged ions were analyzed in reflectron mode. External calibration was performed using a ProteomMass Peptide and protein

MALDI/MS calibration kit (Sigma St. Louis, MO, USA). The three detectable peptide fragments were angiotensin II (human) m/z 1046.542, ACTH fragment 18–39 (human) m/z 2465.198, and insulin-oxidized B chain (bovine) m/z 3494.651. The spectra were obtained through random scanning of the sample surface with an ablation LASER. An average of 100 LASER shots was used to improve the signal-to-noise ratio. Positively charged ions were analyzed with a MALDI-LR in TOF/MS mode instrument (Micromass Waters Corporation, Milford Massachusetts, USA). Three technical replicates from each spectrum were analyzed by MS, and only peptides common to all of the resolved spectra were considered for protein identification.

2.4. Bioinformatics and Statistical Analyses. Chromatogram peaks for each pH fraction (hydrophobicity and p*I*) were converted into band intensity maps using ProteoVue software (Eprogen, Darien, IL, USA). The color intensity of each map was proportional to the relative intensity of each chromatographic peak. Chromatograms obtained from Cd-treated and untreated samples were compared using DeltaVue software (Eprogen, Darien, IL, USA). Protein abundance in treated and untreated samples was semi-quantified by subtracting the difference in peak areas for proteins with the same p*I* and hydrophobicity. A paired *t*-test was used to identify significant differences in abundance of proteins between the control and Cd-treated samples. Only peaks with a ratio >3 or <0.33 between untreated and treated samples were included in the analysis. These threshold levels were chosen as being more conservative than those used in [9].

Peptide mass fingerprinting was carried out using the Mascot program (http://www.matrixscience.com). Proteins were identified by searching through the SWISS-PROT and NCBI nonredundant databases (limited to *Arabidopsis thaliana* and other plant species), as well as against the poplar EST database (http://www.populus.db.umu.se/). The following parameters were used in browsing the protein databases: (i) mass accuracy below 100 ppm, (ii) maximum of one missed cleavage by trypsin, (iii) carbamidomethylation of cysteine, and (iv) oxidation of methionine, where (iii) and (iv) were considered to be fixed modifications. The search was based on the monoisotopic masses of the peptides. For a positive peptide identification, at least five different predicted peptides had to match the observed masses, and the matching peptides had to account for at least 20% of the known sequence.

3. Results and Discussion

3.1. Protein Separation in Leaves. An important aspect of evaluating plants for the phytoremediation of metals is to determine the amount and type of damage to the leaves. An efficient and metal-tolerant photosynthetic system allows plants to maintain high transpiration efficiency, which creates a water flux that can drive metals from the roots and into the stems and leaves where the metal can be compartmentalized. Poplar clone "Poli" was previously selected among 10 Italian poplar clones as having the greatest ability to accumulate Cd in leaves and stem tissues [14].

Several experiments have previously been conducted to characterize the Cd tolerance level of the *Populus nigra* clone "Poli". In one study, leaf samples taken from plants after 21 days of growth in the presence of $50\,\mu M$ $CdSO_4$ showed mild or almost no symptoms of chlorosis and reduction in total leaf area. In comparison, a dose of $50\,\mu M$ $CdSO_4$ halted radical growth and caused symptoms of leaves chlorosis in other clones in this experiment [14]. Other parameters that distinguished "Poli" clones from the other poplar clones included Cd leaf accumulation after 21 days (450 ± 30 mg/kg d.w.) and a high translocation factor. These parameters were previously quantified by atomic absorption spectrophotometer by Zacchini et al. [14].

Given the interesting physiological features of the clone "Poli", proteomic characterization was conducted to further understand its response to Cd. This paper describes how a 2D-LC protein separation system, coupled with MS identification, can be used for comparative plant proteomic analysis. This technique has recently been reported as improving semiquantitative comparisons of protein mixtures [12]. Though it has been widely used in human and bacterial comparative proteomic studies [12, 16, 17], few papers have described its application for plants [13, 18]. One of the advantages of this technique is that a larger amount of protein can be analyzed at once (up to 10 mg) in comparison to the conventional 2D SDS-PAGE (0.1–0.3 mg). One benefit of this system is that it works with crude total soluble extracts. This increases the reproducibility of the protein patterns and avoids the loss of proteins due to sequential purification (e.g., the frequent acid precipitation steps). These aspects are crucial in comparative proteomic studies because they also allow the detection of less abundant proteins.

As leaves from poplar trees have very little protein and an abundance of insoluble substances such as waxes and lignocelluloses that may hamper protein solubilisation, applying a 2D-LC separation technique to plant tissues extracts is an effective method of detecting plant proteins [19]. A standardized procedure for this method has previously been developed by Pirondini et al. [13] to optimize protein solubilization while limiting instrumental interference. Recently, a proteomic approach based on a 2D SDS-PAGE separation system was also applied to poplar leaf extracts to identify several proteins related to the Cd stress response [9]. However, as there are as many as 45 555 putative proteins that could potentially be sequenced from the poplar genome [20] and the proteomic approach has only worked with several hundreds of these proteins to date, the use of different complementary techniques could help to decode the entire proteome of a poplar species. Thus the 2D-LC protein separation technique is a suitable complementary tool that can identify proteins that have been upregulated or downregulated in response to Cd stress.

Using the 2D-LC separation technique, about 600 proteins were resolved from treated ($50\,\mu M$ $CdSO_4$) and untreated ($0\,\mu M$ of $CdSO_4$) poplar cuttings from *Populus*
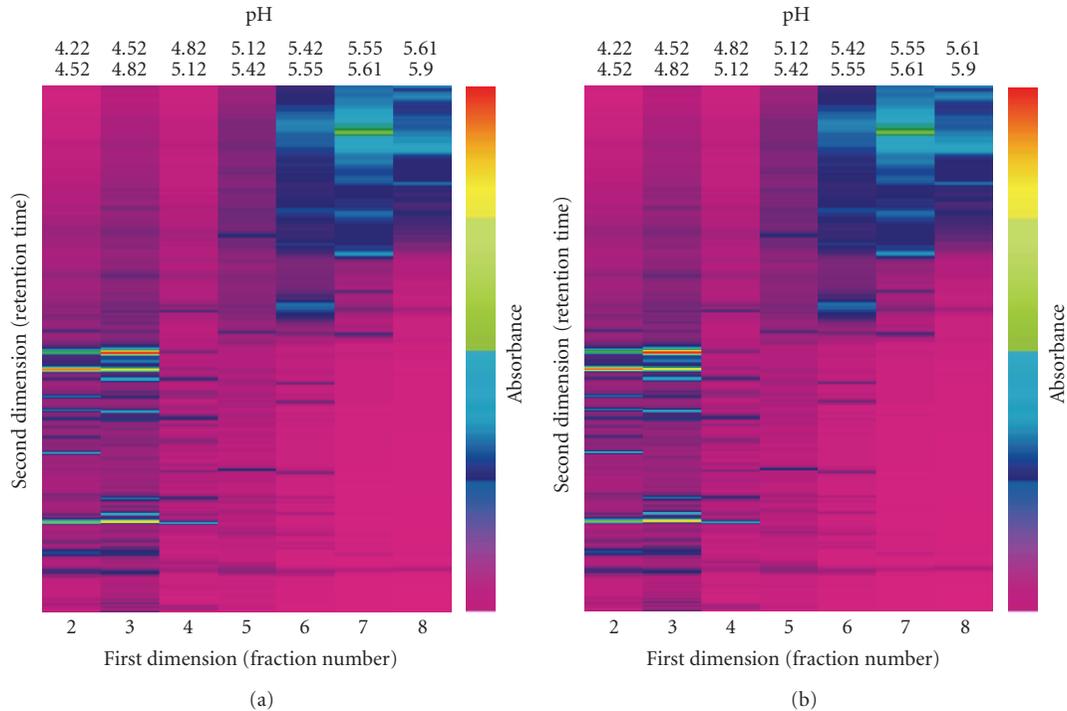
Figure 1: Virtual 2D maps created by the software "ProteoVue" for two independent untreated samples. Proteins were extracted as described in the Materials and Methods Section. Equal amounts of protein (1.3 mg) were loaded into the 1st dimension separation column (HPRP). This figure shows the isoelectric point (pI) fractions (from 4.22 to 5.9) for two independent biological replicates.
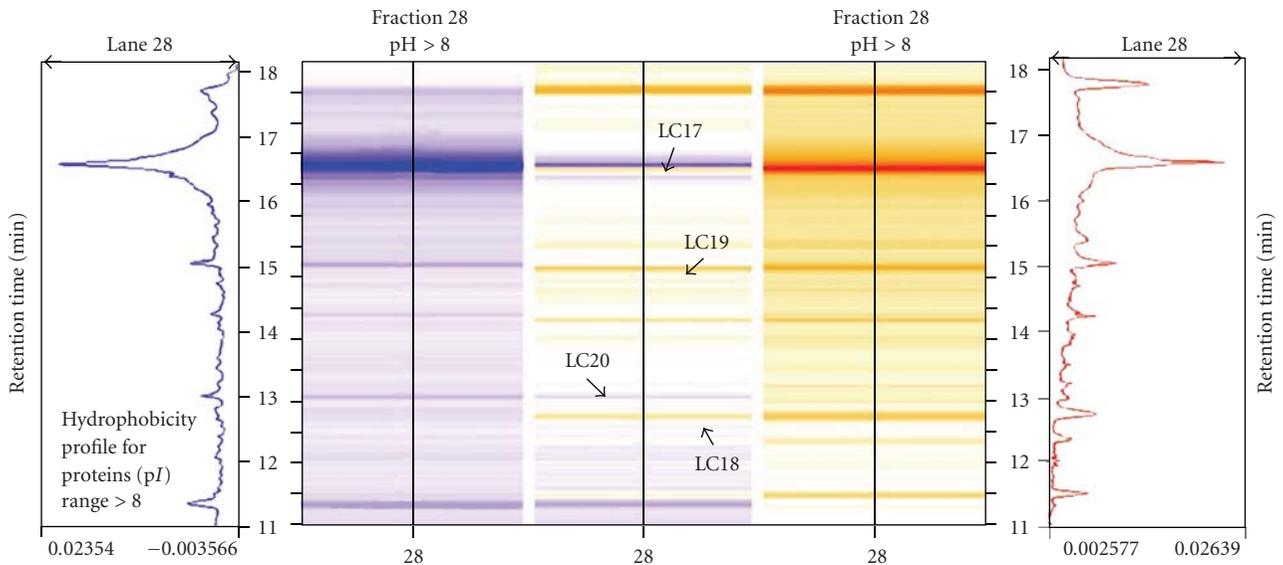
Figure 2: Magnified area of a virtual gel obtained after 2nd dimension separation (HPRC), for the 28th pH fraction, and reproduced by "DeltaVue" software. The protein pattern of untreated plants (0 μM CdSO$_4$) is shown in shades of blue on the left and the protein pattern of treated plants (50 μM CdSO$_4$) is shown in shades of yellow on the right. The center column demonstrates the differences in protein abundance between the control and the treated protein samples, as indicated by arrows at the peaks/ bands for proteins of different abundances.

nigra clone "Poli". Data were displayed on six different virtual gels (three technical replicates for each biological replicate of the control and the treated samples) using ProteoVue software. A comparison of the protein abundance in different samples was carried out using DeltaVue software.

Figure 1 shows the reproducibility of the protein profiles from two technical replicates of the control sample. Thus differences between the treated and untreated samples can be determined by matching the corresponding peaks between samples. This concept is further demonstrated in Figure 2,
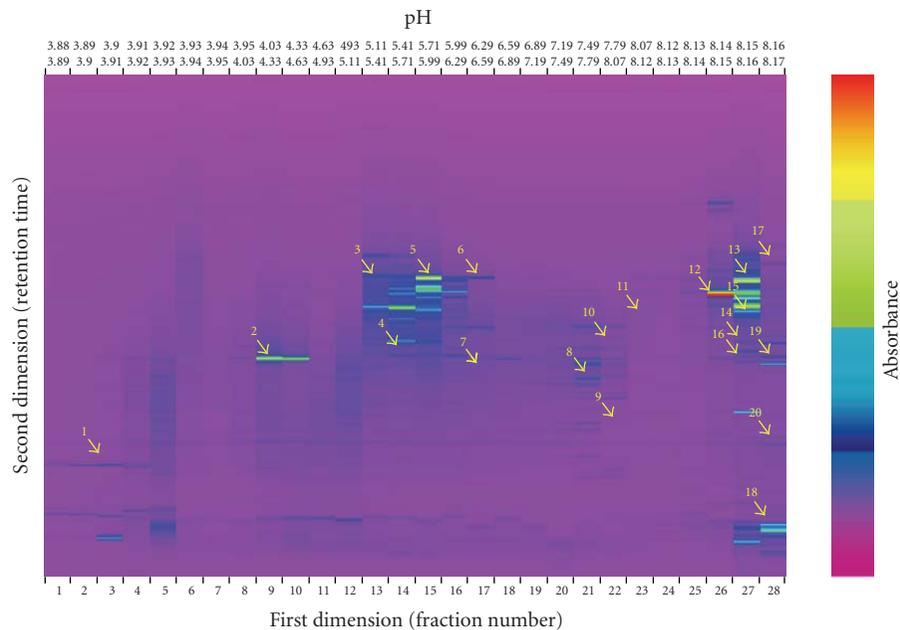
FIGURE 3: A "ProteoVue" 2D map of a leaf protein extract from *Populus nigra* clone "Poli", grown for 3 weeks in nutrient solution supplemented with 50 μM CdSO$_4$. The x-axis is in isoelectric point (pI) units from 4.0 to 8.0. The y-axis displays increasing hydrophobicity. The color scale of the bands represents the relative intensity of each band by UV detection at 214 nm. The proteins detected by MALDI/TOF analysis are numbered and highlighted with yellow arrows.

by highlighting differences in protein abundance in Fraction 28.

Of the peaks that were separated from the two treatments, 126 peaks (13.6%) were three times more or less abundant in either the treated or untreated samples. 50 % of the 126 peaks had an isoelectric point (pI) between 8 and 4, 33% had a pI > 8 (basic), and 17% had a pI < 4 (acidic). Almost half of the 126 identified peaks had an absorbance less than 0.05 at 214 nm, which is under the limit of detection recommended by the manufacturer for MS analysis. Another limitation of all proteins separation techniques is the possibility that there are multiple protein species in a single spot for 2D SDS-PAGE gels, or in a single peak for 2D-LC. Thus each peak eluted out in a single fraction was analyzed by 1D SDS-PAGE, for protein singularity (data not shown). Twenty peaks were clearly identified as single bands on the gel and were further characterized by MS. As the remaining peaks could not clearly be distinguished as individual bands, they were not analyzed any further.

3.2. Protein Identification by Mass Spectrometry. Proteins were identified by eluting out the same peaks from both untreated and Cd-treated samples and then analysing them by MALDI-TOF mass spectrometry. The resulting tryptic peptides were entered into the SWISS-PROT and/or NCBI nonredundant databases. In addition, a BLAST search was performed on the EST poplar database (http://www.populus.db.umu.se/) to confirm the reliability of the results. Results indicated that there was a consistent

correspondence between the experimental and the calculated pI values for most of the analyzed proteins. The few exceptions were likely due to post-transductional modifications of these proteins. Similar inconsistencies have been reported by other authors and have been attributed to the influence of ion concentrations in both the liquid and in the adsorbed phase because a small section of the charged protein appears to bind to the ion-exchange resin [13].

Comparison with the *Arabidopsis thaliana* protein database gave the highest score for almost all twenty proteins. This result confirms the consistency in synteny between *Arabidopsis* and *Populus* for both their genomes and also their proteomes [21]. The proteins identified by MS and their putative function/location are shown in **Figure** 3 and in Tables 1 and 2.

Several of the identified proteins were more abundant in the treated samples than in the untreated samples. For example, Aquaporin NIP-2 is a protein that belongs to the large family of major intrinsic proteins (MIPs) and is present in many eukaryotic organisms from yeast to humans [22, 23]. The physiological role of aquaporins in plants is not well understood, but reports indicate that they may play an important role in transport selectivity and gating (i.e. opening and closure of pores) and can help transport or diffuse physiologically important molecules such as CO_2, H_2O_2, ammonium (NH_4^+), and its gaseous conjugated base (NH_3, ammonia) through the cell membrane. Aquaporins can also aid in the uptake of metalloids such as As and Sb and are involved in carbon fixation mechanisms, cell signaling transduction, stress responses, and cell osmotic and pH stabilization [24]. Another protein identified as

Table 1: Proteins regulated by Cd treatment in leaves of *Populus nigra* clone "Poli", separated by 2D-LC.

Nos.[a]	Homologous protein[b]	Accession number[b]	Measured pI interval[c]	Predicted pI/Mr (Da)[c]	Score / coverage (%)[d]	Ratio[e]	ESTs poplar[f]
LC1	Aquaporin	NIP12_ARATH	<4.00	8.63/31249	56/40	+5.6	P047H06
LC2	Dehydration-responsive protein	gi|18405331	4.03–4.33	8.60/67472	60/35	+3.4	P082D06
LC3	Predicted protein	gi|162695261	5.11–5.41	6.20/45886	69/32	−7.2	T020D11
LC4	Acetyl coenzime A carboxylase	gi|74272305	5.41–5.71	5.60/120177	69/35	−3.5	Q001C04
LC5	NADPH isocitrate dehydrogenase	gi|169989	5.71–5.99	6.13/49508	51/28	+4.4	N006H08
LC6	Thylakoidal processing peptidase 2	TPP2_ARATH	6.29–6.59	6.54/40581	56/21	+4.8	T040C06
LC7	Hypothetical protein	gi|125603717	6.29–6.59	6.67/47148	71/26	−3.0	Y004E07
LC8	Predicted protein	gi|168009327	7.49–7.79	7.22/42542	69/40	+5.3	V017A06
LC9	Chaperone ClpB	CLPB_RHOBA	>8.00	5.06/97498	110/56	+8.4	P058E06
LC10	Rubisco large subunit	gi|33636375	>8.00	5.13/6207	69/80	−8.5	No match
LC11	Aldo/keto reductase protein	gi|15232354	>8.00	6.46/35185	64/35	+4.5	V056F06
LC12	Cytochrome P450	C71BT_ARATH	>8.00	8.06/56318	105/64	+7.3	M122H12
LC13	ATP-dependent Clp protein	CLPR1_ARATH	>8.00	8.81/42942	51/35	+4.7	UM61TE04
LC14	Cysteine proteinase inhibitor 2	CYT2_ORYSJ	>8.00	8.82/16550	56/25	+3.2	FL_GENBANK_63
LC15	Hydroxyacylglutathione hydrolase 1	GLO2M_ARATH	>8.00	8.83/36933	51/40	−4.6	S065B02
LC16	Hypothetical protein	gi|38175624	>8.00	9.10/27188	69/36	−4.5	N054D02
LC17	GTP binding protein	gi|145359760	>8.00	9.46/20079	60/40	−3.2	K027P68
LC18	Protein kinase APK1B	APK1B_ARATH	>8.00	9.37/45946	56/37	+5.3	N041C12
LC19	CRS2A chloroplastic intron splicing facilitator	CRS2A_ARATH	>8.00	9.40/27356	69/60	+3.5	K028P24
LC20	Nucleotide translocator	gi|444790	>8.00	9.78/43995	92/43	−7.4	T027D03

[a]The nos. refer to the spot numbers as given in Figure 3. [b]Putative protein identification and accession number for the closest match in the databases. [c]pI values as estimated from the 2D-LC together with the predicted pI and Mr of the closest match in the SWISS PROT and NCBI databases. [c]pI values as estimated from the 2D-LC together with the predicted pI and Mr of the closest match in the databases. [d]Score and percentage of coverage of the matching peptide sequence tags. [e]Ratio between values of the peak areas of the proteins from treated leaves (50 μM CdSO$_4$) and untreated leaves (0 μM CdSO$_4$) obtained from the Deltavue software. [f]Matches with the poplar EST database.

TABLE 2: Functional classes and putative locations of the proteins identified by MS.

Protein class	Protein name	Localization	Predicted function or domains
Uptake	Aquaporin NIP-2	Plasma membrane	Channel/ pore transporter
Proteolysis	Thylakoidal processing peptidase	Chloroplast membrane	Peptidase activity
Stress proteins	Cysteine proteinase inhibitor	Cytosol	Inhibition of exogenous proteases such as those present in digestive tracks of insects and nematodes
	Aldo/keto reductase	Cytosol	Oxidoreductase activity involved in response to cadmium ions
	Cytochrome P450	Endoplasmic reticulum	Ferulate 5-hydrolase activity, monooxygenase activity
	Dehydration responsive protein	Golgi apparatus	Biological process unknown
	ClpB, chloroplastic	Chloroplast	ATP binding, ATPase activity
	ATP-dependent Clp	Chloroplast	Protease activity
Metabolism	NADPH isocitrate dehydrogenase	Mitochondrion	TCA cycle
	CRS2A, chloroplastic intron splicing facilitator	Chloroplast	Aminoacyl-tRNA hydrolase activity
	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	Chloroplast	Calvin cycle
	Nucleotide translocator	Mitochondrial inner membrane	ATP : ADP antiporter activity
	Acetyl-CoA carboxylase	Mitochondrion	Fatty acid metabolism
	Predicted protein	Mitochondrion	Succinyl-CoA synthetase subunit beta
Regulatory protein	APK1B kinase	Chloroplast	Serine/ threonine activity
	GTP binding protein	Mitochondrion	Protein targeting, sorting and translocation
	Predicted protein	—	Serine/ threonine kinase domain
Glutathione pathway	Hydroxyacylglutathione hydrolase 1	Mitochondrion	Thiolesterase that catalyzes the hydrolysis of S-D-lactoyl glutathione to form glutathione and D-lactic acid

Predictions were obtained by checking putative protein acquisition with the annotation in SWISS-PROT, NCBI, and databases.

being particularly abundant in the Cd-treated samples was the thylakoidal processing peptidase, which is a chloroplast enzyme relevant for protein maturation within organelles [25].

Kinase 1B (APK1B), a protein homolog in Arabidopsis, was also found to be more abundant in leaves of Cd-treated plants. It is involved in phosphorylation of tyrosine, serine, and threonine residues and contributes to signal transduction in the chloroplast [26]. In animal cells, protein tyrosine kinases have been identified in both receptor and cytosolic forms and play a role in cell growth, development and the oxidative stress response [27].

Two proteins involved in the stress response, ClpB and ATP-dependent Clp, were more prevalent in the Cd-treated samples. The heat shock protein ClpB is a member of the caseinolytic protease (Clp) family, whose members act as molecular chaperones in many organisms [28, 29]. ClpB is found in the chloroplastic stroma and helps with thermotolerance. In *Arabidopsis thaliana*, ClpB-p knockouts mutants did not survive for long as they failed to accumulate chlorophyll and could not properly develop the chloroplasts. Thus, in plants, ClpB proteins are not limited to the heat stress response, but play an essential role in chloroplast

development under different stress conditions (i.e., exposure to toxic metals) [30]. The related protein, ATP-dependent Clp, is also involved in chloroplast functions essential for cell survival. The recent characterization of several chloroplast proteases determined that ClpRs are the noncatalytic subunits of the Clp core proteolytic complex [31].

The chloroplast-specific NADPH isocitrate dehydrogenase was also more abundant in the Cd-treated samples, along with CRS2, a putative chloroplast RNA Splicing 2 factor that is required for splicing of group II introns. The importance of RNA splicing, and alternative RNA splicing patterns for plant survival in response to stress conditions, has been deduced using a transient expression system in monocots [32]. In plant chloroplasts and in animal mitochondria, different metabolic reactions are coupled with NADPH, including those involved in thiol group reduction. Mitochondrial NADP$^+$-dependent isocitrate dehydrogenase is an enzyme linked to the above mentioned process, although the precise role of this enzyme is still unknown. It was suggested that this protein may be a regulatory switch in the tricarboxylic acid cycle flux and in the reductive modulation of alternative oxidase (AOX) [33]. Castro-Guerrero et al. [34] have demonstrated that an enhancement

in AOX appears to be related to the Cd-resistance mechanism developed in the algae *E. gracilis*. When their primary respiration pathway is impaired, the algae switch to an alternative oxidative route to maintain a relatively unaltered level of metabolism [34].

Interestingly, a cysteine proteinase inhibitor was more abundant in the Cd–treated leaves. This enzyme can potentially improve resistance to biotic stresses such as pathogens and insects in both animals and plants [35, 36]. An aldoketoreductase protein, previously found in animals and recently identified in plants in response to stress, was also more abundant after Cd stress. Oberschall et al. [37] have shown that a novel plant NADPH-dependent aldo-keto reductase functions in a detoxification pathway which can prevent stress damage.

Another protein more abundant in the Cd-treated plants is a member of the cytochrome P450 family, which have a key role in the response to abiotic and biotic stresses both in animals and plants [38, 39]. Information on the mechanism of how metal ions inhibit NADPH-cytochrome P450 reductase is limited in comparison with the effect of organic pollutants on both poplars and *Arabidopsis* [40]. The cytochrome P450 enzymes also regulate the production of the plant hormone abscisic acid (ABA). As ABA 8′-hydroxylase belongs to a cytochrome P450 family, it implies that this protein plays a major regulatory role in controlling the level of ABA in plants [41]. ABA is involved in apoptosis, senescence, and leaf shedding, and ABA regulation is essential for maintenance of plant fitness in abiotic stress conditions, such as drought and also Cd stress [42].

Eight proteins were identified as being less abundant after the Cd treatment. A RuBisCo large subunit was found to be more abundant in the untreated samples, as the photosynthetic system was likely damaged or less efficient in the Cd-treated plants. In this case, the measured and the theoretical p*I* values were different, but the "score" and "coverage" parameters of the protein were high enough to confirm its identification. In literature, examples of post-transductional modifications to the large and small subunits of RuBisCo such as acetylations, deformylation, methylation, and phosphorylation support the possibility of different p*I*s [43].

Four mitochondrial proteins were found to be less abundant after the Cd treatment. (1) Hydroxyacylglutathione hydrolase 1 is a mitochondrial precursor that functions as a thiolesterase and catalyzes the hydrolysis of S-D-lactoylglutathione to form glutathione and D-lactic acid. This enzyme has been described in animals and in plants [44] and the ability of glutathione to bind Cd has been well established [45]. (2) Acetyl-CoA carboxylase is an enzyme that catalyzes the first step in fatty acid synthesis, the carboxylation of acetyl-CoA to malonyl-CoA. Two physically distinct types of this enzyme are present in nature, and most plants have both forms. The heteromeric form (four subunits similar to those found in prokaryotes) is found in the plastids, where de novo synthesis of fatty acids occurs. The homomeric form (single large polypeptide similar to that found in eukaryotes) is found in the cytosol. This enzyme may play a role in the oxidative stress response, (i.e.,

after exposure to metals) [46]. (3) A GTP-binding protein. G proteins bind to heterotrimeric GTP-binding proteins and transduce extracellular signals into intracellular signals by activating effector molecules (i.e., adenylate cyclases), which catalyze cAMP formation. cAMP signaling pathways have been reported as being able to protect cells from ionizing radiation-induced apoptosis. However the mechanism of this protection is still unknown [47]. (4) A nucleotide translocator is protein known to be involved in the alteration of mitochondrial membrane permeability for free radicals and metal ions in animals and plants [48, 49].

Another paper was recently published on proteomic responses in poplar leaves after exposure to a prolonged Cd treatment [9]. The papers differed in (i) plant material: *Populus nigra versus Populus tremula*, (ii) length of the Cd treatment and concentration: $50\,\mu M$ Cd for 21 days *versus* $20\,\mu M$ Cd for 28 days, (iii) protein extraction, separation and identification procedures: 2D-LC *versus* 2D-PAGE gel, and (iv) threshold level for the comparative analysis: 3-fold *versus* 1.5-fold ratios. However, some common features were found with respect to the classes and the cellular location of some of the identified proteins, including stress proteins, carbohydrate and general metabolism-related proteins, and proteolysis proteins: Table 2 and [9]. Kieffer et al. [9] concluded that after the first "alarm phase" there was a stabilization of the stress response and plants remained photosynthetically active. In addition, plants increased their energy supply by activating mitochondrial respiration. In comparison, this paper determined that some regulatory proteins such as homolog of a GTP-binding protein and the protein kinase APK1B that were more abundant after Cd treatment (21 days) were not detected in the *Populus tremula* clone [9]. It was hypothesized that these proteins may account for the robust response by Poli clone to Cd stress. The synthesis of these proteins could limit the damages and/or optimize the different pathways useful for Cd adaptation and Cd homeostasis (Table 2). To date, no studies have been found regarding the regulation of these proteins under Cd stress, probably due to their low abundance in plant samples, confirming the value of this separation technique.

4. Conclusion

Environmental exposure to toxic elements such as Cd forces both plants and animals to enter into a "tradeoff" situation to minimize the impact of these compounds on the organism. "Tradeoffs" may include the activation of a stress defence response, reduction of metabolic rate, and/or activation of novel metabolic routes. A phenotypic study of the adaptive response is technically more difficult than transcriptomic analysis. For instance, the proteomic analyses described in this study and in Kieffer et al. by [9] focus on only a small percentage of the putative poplar proteomes. However, phenotypic plasticity and environmental adaptations from "tradeoffs" between gene expression and environmental conditions would be difficult to quantify without protein analysis. Comparative proteomics of metal treated and

untreated plants (e.g., Poli) is likely to be applicable to multiple areas of phytoremediation in future. In particular, proteins with altered synthesis due to changes in gene expression may be used to design molecular markers for the selection of more advantageous genotypes.

Acknowledgments

References

[1] S. Dudka and W. P. Miller, "Accumulation of potentially toxic elements in plants and their transfer to human food chain," *Journal of Environmental Science and Health B*, vol. 34, no. 4, pp. 681–708, 1999.

[2] A. Schützendübel and A. Polle, "Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization," *Journal of Experimental Botany*, vol. 53, no. 372, pp. 1351–1365, 2002.

[3] B. H. Robinson, T. M. Mills, D. Petit, L. E. Fung, S. R. Green, and B. E. Clothier, "Natural and induced cadmium-accumulation in poplar and willow: implications for phytoremediation," *Plant and Soil*, vol. 227, no. 1-2, pp. 301–306, 2000.

[4] R. Unterbrunner, M. Puschenreiter, P. Sommer, et al., "Heavy metal accumulation in trees growing on contaminated sites in Central Europe," *Environmental Pollution*, vol. 148, no. 1, pp. 107–114, 2007.

[5] N. Marmiroli and S. C. McCutcheon, "Making phytoremediation a successful technology," in *Phytoremediation: Transformation and Control of Contaminants*, S. C. McCutcheon and J. L. Schnoor, Eds., pp. 75–107, Wiley-Interscience, Hoboken, NJ, USA, 2003.

[6] J. E. van de Mortel, H. Schat, P. D. Moerland, et al., "Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in *Arabidopsis thaliana* and the related Zn/Cd-hyperaccumulator *Thlaspi caerulescens*," *Plant, Cell and Environment*, vol. 31, no. 3, pp. 301–324, 2008.

[7] M. Rossignol, J.-B. Peltier, H.-P. Mock, A. Matros, A. M. Maldonado, and J. V. Jorrín, "Plant proteome analysis: a 2004–2006 update," *Proteomics*, vol. 6, no. 20, pp. 5529–5548, 2006.

[8] J.-E. Sarry, L. Kuhn, C. Ducruix, et al., "The early responses of *Arabidopsis thaliana* cells to cadmium exposure explored by protein and metabolite profiling analyses," *Proteomics*, vol. 6, no. 7, pp. 2180–2198, 2006.

[9] P. Kieffer, P. Schroeder, J. Dommes, L. Hoffmann, J. Renaut, and J.-F. Hausman, "Proteomic and enzymatic response of poplar to cadmium stress," *Journal of Proteomics*, vol. 72, no. 3, pp. 379–396, 2009.

[10] J. K. C. Rose, S. Bashir, J. J. Giovannoni, M. M. Jahn, and R. S. Saravanan, "Tackling the plant proteome: practical approaches, hurdles and experimental tools," *The Plant Journal*, vol. 39, no. 5, pp. 715–733, 2004.

[11] T. Rabilloud, "Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains," *Proteomics*, vol. 2, no. 1, pp. 3–10, 2002.

[12] J.-P. Lambert, M. Ethier, J. C. Smith, and D. Figeys, "Proteomics: from gel based to gel free," *Analytical Chemistry*, vol. 77, no. 12, pp. 3771–3787, 2005.

[13] A. Pirondini, G. Visioli, A. Malcevschi, and N. Marmiroli, "A 2-D liquid-phase chromatography for proteomic analysis in plant tissues," *Journal of Chromatography B*, vol. 833, no. 1, pp. 91–100, 2006.

[14] M. Zacchini, F. Pietrini, G. Scarascia Mugnozza, V. Iori, L. Pietrosanti, and A. Massacci, "Metal tolerance, accumulation and translocation in poplar and willow clones treated with cadmium in hydroponics," *Water, Air, and Soil Pollution*, vol. 197, no. 1–4, pp. 23–34, 2009.

[15] D. I. Arnon and R. D. Hoagland, "Crop production in artificial culture solutions and in soils with special reference to factors influencing yield and absorption of inorganic nutrient," *Soil Science*, vol. 50, pp. 463–483, 1940.

[16] V. Ruelle, N. Falisse-Poirrier, B. ElMoualij, et al., "An immuno-PF2D-MS/MS proteomic approach for bacterial antigenic characterization: to *Bacillus* and beyond," *Journal of Proteome Research*, vol. 6, no. 6, pp. 2168–2175, 2007.

[17] J. D. Schlautman, W. Rozek, R. Stetler, R. L. Mosley, H. E. Gendelman, and P. Ciborowski, "Multidimensional protein fractionation using ProteomeLab PF 2D^{TM} for profiling amyotrophic lateral sclerosis immunity: a preliminary report," *Proteome Science*, vol. 6, article 26, 2008.

[18] S. Komatsu, X. Zang, and N. Tanaka, "Comparison of two proteomics techniques used to identify proteins regulated say gibberellin in rice," *Journal of Proteome Research*, vol. 5, no. 2, pp. 270–276, 2006.

[19] C. Plomion and C. Lalanne, "Protein extraction from woody plants," in *Plant Proteomics: Methods and Protocols*, H. Thiellement, M. Zivy, C. Damerval, and V. Méchin, Eds., vol. 335, Humana Press, Totowa, NJ, USA, 2007.

[20] G. A. Tuskan, S. DiFazio, S. Jansson, et al., "The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)," *Science*, vol. 313, no. 5793, pp. 1596–1604, 2006.

[21] G. Taylor, "*Populus*: Arabidopsis for forestry. Do we need a model tree?" *Annals of Botany*, vol. 90, no. 6, pp. 681–689, 2002.

[22] A. Tanghe, P. Van Dijck, F. Dumortier, A. Teunissen, S. Hohmann, and J. M. Thevelein, "Aquaporin expression correlates with freeze tolerance in baker's yeast, and overexpression improves freeze tolerance in industrial strains," *Applied and Environmental Microbiology*, vol. 68, no. 12, pp. 5981–5989, 2002.

[23] B. Yang, "The human aquaporin gene family," *Current Genomics*, vol. 1, no. 1, pp. 91–102, 2000.

[24] G. P. Bienert, M. Thorsen, M. D. Schüssler, et al., "A subgroup of plant aquaporins facilitate the bi-directional diffusion of

As(OH)$_3$ and Sb(OH)$_3$ across membranes," *BMC Biology*, vol. 6, article 26, 2008.

[25] B. K. Chaal, K.-I. Ishida, and B. R. Green, "A thylakoidal processing peptidase from the heterokont alga *Heterosigma akashiwo*," *Plant Molecular Biology*, vol. 52, no. 2, pp. 463–472, 2003.

[26] T. Hirayama and A. Oka, "Novel protein kinase of *Arabidopsis thaliana* (APK1) that phosphorylates tyrosine, serine and threonine," *Plant Molecular Biology*, vol. 20, no. 4, pp. 653–662, 1992.

[27] S. K. Hanks, A. M. Quinn, and T. Hunter, "The protein kinase family: conserved features and deduced phylogeny of the catalytic domains," *Science*, vol. 241, no. 4861, pp. 42–52, 1988.

[28] M. Kitagawa, C. Wada, S. Yoshioka, and T. Yura, "Expression of ClpB, an analog of the ATP-dependent protease regulatory subunit in *Escherichia coli*, is controlled by a heat shock sigma factor (sigma 32)," *Journal of Bacteriology*, vol. 173, no. 14, pp. 4247–4253, 1991.

[29] D. A. Parsell, A. S. Kowal, and S. Lindquist, "*Saccharomyces cerevisiae* Hsp104 protein. Purification and characterization of ATP-induced structural changes," *Journal of Biological Chemistry*, vol. 269, no. 6, pp. 4480–4487, 1994.

[30] U. Lee, I. Rioflorido, S.-W. Hong, J. Larkindale, E. R. Waters, and E. Vierling, "The Arabidopsis ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development," *The Plant Journal*, vol. 49, no. 1, pp. 115–127, 2007.

[31] Z. Adam, A. Rudella, and K. J. van Wijk, "Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts," *Current Opinion in Plant Biology*, vol. 9, no. 3, pp. 234–240, 2006.

[32] R. M. Sinibaldi and I. J. Mettler, "Intron splicing and intron-mediated enhanced expression in monocots," *Progress in Nucleic Acid Research and Molecular Biology*, vol. 42, pp. 229–257, 1992.

[33] G. R. Gray, A. R. Villarimo, C. L. Whitehead, and L. McIntosh, "Transgenic tobacco (*Nicotiana tabacum* L.) plants with increased expression levels of mitochondrial NADP$^+$-dependent isocitrate dehydrogenase: evidence implicating this enzyme in the redox activation of the alternative oxidase," *Plant and Cell Physiology*, vol. 45, no. 10, pp. 1413–1425, 2004.

[34] N. A. Castro-Guerrero, J. S. Rodríguez-Zavala, A. Marín-Hernández, S. Rodríguez-Enríquez, and R. Moreno-Sánchez, "Enhanced alternative oxidase and antioxidant enzymes under Cd^{2+} stress in Euglena," *Journal of Bioenergetics and Biomembranes*, vol. 40, no. 3, pp. 227–235, 2008.

[35] S. A. Masoud, L. B. Johnson, F. F. White, and G. R. Reeck, "Expression of a cysteine proteinase inhibitor (oryzacystatin-I) in transgenic tobacco plants," *Plant Molecular Biology*, vol. 21, no. 4, pp. 655–663, 1993.

[36] R. Ettari, N. Micale, T. Schirmeister, et al., "Novel peptidomimetics containing a vinyl ester moiety as highly potent and selective falcipain-2 inhibitors," *Journal of Medicinal Chemistry*, vol. 52, no. 7, pp. 2157–2160, 2009.

[37] A. Oberschall, M. Deak, K. Torok, et al., "A novel aldose/aldehyde reductase protects transgenic plants against lipid peroxidation under chemical and drought stresses," *The Plant Journal*, vol. 24, no. 4, pp. 437–446, 2000.

[38] W. M. Lee, "Drug-induced hepatotoxicity," *The New England Journal of Medicine*, vol. 349, no. 5, pp. 474–485, 2003.

[39] Y. Narusaka, M. Narusaka, M. Seki, et al., "Crosstalk in the responses to abiotic and biotic stresses in Arabidopsis: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray," *Plant Molecular Biology*, vol. 55, no. 3, pp. 327–342, 2004.

[40] S. L. Doty, C. A. James, A. L. Moore, et al., "Enhanced phytoremediation of volatile environmental pollutants with transgenic trees," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 43, pp. 16816–16821, 2007.

[41] S. A. Quarrie, M. Gulli, C. Calestani, A. Steed, and N. Marmiroli, "Location of a gene regulating drought-induced abscisic acid production on the long arm of chromosome 5A of wheat," *Theoretical and Applied Genetics*, vol. 89, no. 6, p. 794, 1994.

[42] Y. T. Hsu and C. H. Kao, "Role of abscisic acid in cadmium tolerance of rice (*Oryza sativa* L.) seedlings," *Plant, Cell and Environment*, vol. 26, no. 6, pp. 867–874, 2003.

[43] R. L. Houtz, R. Magnani, N. R. Nayak, and L. M. A. Dirk, "Co- and post-translational modifications in Rubisco: unanswered questions," *Journal of Experimental Botany*, vol. 59, no. 7, pp. 1635–1645, 2008.

[44] T. H. Hutchinson, Y. Mahshid, R. Jönsson, C. Björklund, and K. Kenne, "Proteomic analysis of phospholipidosis in citalopram treated U937 cells—support for the cholesterol biosynthesis hypothesis," *Toxicology in Vitro*, vol. 22, no. 5, pp. 1198–1204, 2008.

[45] D. G. Mendoza-Cózatl, E. Butko, F. Springer, et al., "Identification of high levels of phytochelatins, glutathione and cadmium in the phloem sap of *Brassica napus*. A role for thiol-peptides in the long-distance transport of cadmium and the effect of cadmium on iron translocation," *The Plant Journal*, vol. 54, no. 2, pp. 249–259, 2008.

[46] Y. Sasaki and Y. Nagano, "Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding," *Bioscience, Biotechnology and Biochemistry*, vol. 68, no. 6, pp. 1175–1184, 2004.

[47] S.-Y. Kim, M. Seo, J.-M. Oh, E.-A. Cho, and Y.-S. Juhnn, "Inhibition of gamma ray-induced apoptosis by stimulatory heterotrimeric GTP binding protein involves Bcl-xL down-regulation in SH-SY5Y human neuroblastoma cells," *Experimental and Molecular Medicine*, vol. 39, no. 5, pp. 583–593, 2007.

[48] J. J. Chen, H. Bertrand, and B. P. Yu, "Inhibition of adenine nucleotide translocator by lipid peroxidation products," *Free Radical Biology and Medicine*, vol. 19, no. 5, pp. 583–590, 1995.

[49] H. Hashimoto, R. Nishi, M. Umeda, H. Uchimiya, and A. Kato, "Isolation and characterization of a rice cDNA clone encoding ATP/ADP translocator," *Plant Molecular Biology*, vol. 22, no. 1, pp. 163–164, 1993.

Research Article

Proteomics Strategy for Identifying Candidate Bioactive Proteins in Complex Mixtures: Application to the Platelet Releasate

Roisin O'Connor,[1] Lorna M. Cryan,[2] Kieran Wynne,[1] Andreas de Stefani,[1] Desmond Fitzgerald,[2] Colm O'Brien,[2] and Gerard Cagney[1]

[1] School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin 4, Ireland
[2] School of Medicine and Medical Science, Conway Institute, University College Dublin, Dublin 4, Ireland

Correspondence should be addressed to Gerard Cagney, gcagney@gmail.com

Proteomic approaches have proven powerful at identifying large numbers of proteins, but there are fewer reports of functional characterization of proteins in biological tissues. Here, we describe an experimental approach that fractionates proteins released from human platelets, linking bioassay activity to identity. We used consecutive orthogonal separation platforms to ensure sensitive detection: (a) ion-exchange of intact proteins, (b) SDS-PAGE separation of ion-exchange fractions and (c) HPLC separation of tryptic digests coupled to electrospray tandem mass spectrometry. Migration of THP-1 monocytes in response to complete or fractionated platelet releasate was assessed and located to just one of the forty-nine ion-exchange fractions. Over 300 proteins were identified in the releasate, with a wide range of annotated biophysical and biochemical properties, in particular platelet activation, adhesion, and wound healing. The presence of PEDF and involucrin, two proteins not previously reported in platelet releasate, was confirmed by western blotting. Proteins identified within the fraction with monocyte promigratory activity and not in other inactive fractions included vimentin, PEDF, and TIMP-1. We conclude that this analytical platform is effective for the characterization of complex bioactive samples.

1. Introduction

Platelets are anucleate cells that are important for haemostasis, thrombosis, and atherosclerotic disease. A number of bleeding disorders arise as a result of mutations in the genes for proteins involved in platelet aggregation. Thus, altered or deregulated platelet function underpins many diseases, and platelet proteins are potential targets for novel therapeutic agents. Previous proteomic studies of intact platelets have collectively identified hundreds of proteins using a variety of fractionation strategies including 2-dimensional electrophoresis (2DE), multidimensional chromatographic separations, membrane prefractionation techniques, and adsorption to combinatorial hexapeptide ligand libraries [1–6]. Following activation by agonists such as thrombin, platelets release storage granules and membrane vesicles that contain prothrombotic (e.g., fibrinogen), mitogenic (e.g., platelet derived growth factor), immunomodulatory (e.g., neutrophil-activating peptide 2), and adhesive (e.g., platelet endothelial cell adhesion molecule) proteins. A previous study from our laboratory using a MuDPIT (multidimensional protein identification technology) approach identified over 300 proteins secreted by platelets upon thrombin activation [1]. These proteins may modulate the interaction of platelets with their local cellular environment. Indeed, platelet releasate has previously been shown to induce endothelial cell permeability, endothelial cell chemotaxis, and corneal epithelial cell proliferation in cellular assays [7–9].

The issue of abundant (often housekeeping) proteins masking regulatory proteins of lower abundance (such as signaling proteins and cytokines) continues to be a challenging issue for proteomics particularly in the case of biofluids. Plasma has a significant dynamic range, with more than 10 orders of magnitude separating albumin concentration and the rarest measurable proteins identified to date [10]. Protein and peptide chromatography prior to MS analysis can partly address this issue. The classical example of this

is MuDPIT, first described by Andrew Link and colleagues [11]. In the first dimension, ion-exchange chromatography (IEC) separates peptides based on ionic interactions with the solid phase and an increasing salt buffer. This is coupled to RP-HPLC to create a second dimension of separation prior to MS analysis. Complex biological samples have also been successfully separated in multiple dimensions at the protein level. For instance, a number of studies have used 1D SDS-PAGE to separate proteins by molecular weight prior to MS analysis [12–15].

While modern proteomics experiments permit the analysis of hundreds to thousands of proteins in complex samples, the most powerful use of this data would combine information on protein activity with the identities of the active proteins. The emerging field of chemical proteomics [16] addresses this question elegantly for an increasing number of enzymatic functions through formation of covalent enzyme-substrate conjugates [17]. However, many higher-level cellular functions encompassing multiple steps are not amenable to these approaches. A limited number of studies involving proteomic fractionation combined with biological assay and subsequent MS identification of active proteins have been reported [18–20]. Here we use IEC to separate the platelet releasate for subsequent assay of monocyte migratory activity. IEC relies on electrostatic interactions between the proteins and the chromatography matrix, so that fractionation primarily depends on net protein charge. We chose a scheme, incorporating both cation and anion resins, in preference to a size-based fractionation scheme because many secreted proteins are small (<20 KDa) and so would cofractionate. We further separated the IEC fractions by SDS-PAGE, carrying out multiple (~30) LCMS runs per IEC fraction, thus ensuring extensive coverage of the platelet releasate proteome. By correlating the presence of proteins in active fractions (and conversely by discounting the contribution of proteins detected in inactive fractions), a more focused list of active proteins could be obtained compared to those obtained from conventional proteomics approaches.

2. Methods

2.1. Preparation of Platelet Releasate.
Plasma and washed platelet samples were prepared from 100 mL of blood drawn from healthy human volunteers free from analgesic medication for 10 days using a modification of the method described by Mendelsohn [21]. Following centrifugation of blood for 10 minutes at 150 g, platelet-rich plasma (PRP) was harvested. Platelets were pelleted by centrifugation at 720 g for 10 minutes and subsequently resuspended in a modified HEPES buffer (JNL buffer) (130 mM NaCl, 10 mM trisodium citrate, 9 mM NaHCO$_3$, 6 mM dextrose, 0.9 mM MgCl$_2$, 0.81 mM KH$_2$PO$_4$, 10 mM Tris pH 7.4) at a concentration of 2×10^9 platelets/mL or 2×10^8 platelets/mL.

Platelet suspensions were supplemented with 1.8 mM CaCl$_2$, incubated at 37°C in an aggregometer under constant stirring conditions (1100 rev/min), and stimulated with 5 μM thrombin receptor-activating peptide (TRAP) for 3 minutes. Platelet releasate was isolated as previously described [1]. Briefly, platelets were removed from the releasate by two sequential centrifugal spins at 750 g, protease inhibitor cocktail (Merck Biosciences, UK) was added, and releasate samples were stored at −20°C.

2.2. Ion-Exchange Chromatography.
Platelet releasate samples (from 2×10^9 platelets/mL preparations) were desalted using Zeba spin columns (Pierce, Rockford, Illinois) as per manufacturer instructions prior to ion-exchange chromatography (IEC) fractionation. Subsequently, 0.75 mL of desalted platelet releasate was injected onto ProPac anion and cation-exchange columns (Dionex, UK) configured in series on the Ultimate TM 3000 LC system (Dionex, UK). The flow rate was set at 200 μL/min and fractions were collected every minute into a 96-well plate. A UV detector (wavelength of 280 nm) was used to identify eluting peptides. After an initial 10-minutes equilibration, a gradient of 0.02 M–0.5 M salt was delivered to the columns by appropriate mixing of IEC buffer A (20 mM K$_2$HPO$_4$ pH 9.0) and IEC buffer B (20 mM K$_2$HPO$_4$ pH 9.0, 1 M NaCl) over a time course of 25-minutes. This was followed by a 5-minute wash with IEC buffer B ("clean off") and 5-minute equilibration with IEC buffer A.

2.3. Cell Culture and Migration Assay.
A monocytic leukaemia cell line, THP-1, was obtained from EACC (Salisbury, UK). Cells were grown in RPMI 1640 medium, supplemented with 10% FBS, 2 mM/L L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin. The cells were passaged every 5–7 days. All cells were maintained as a suspension culture at 37°C in 5% CO$_2$/95% ambient air. THP-1 migration was measured using a Boyden chamber assay. THP-1 cells were centrifuged at 1000 g and resuspended in serum-free media (SFM) at a concentration of 1×10^6 cells/mL. Of this cell suspension, 100 μL was added to the upper chamber of an 8 μm pore-size transwell (Corning BV, Schiphol-Rijk, The Netherlands). For the purpose of the migration assay, 2 μL of three successive IEC fractions (from 1–49) were pooled and added to the lower chamber. SFM was added to a final volume of 1 mL. A positive control of 10% FBS in SFM was used. The transwell inserts were placed into the lower chamber and incubated for 2 hours at 37°C in 5% CO$_2$/95% ambient air. The number of two populations of cells migrating to the bottom chamber was assessed, the cells migrating to the bottom chamber and adhering to the membrane and the cells migrating to the bottom chamber without adhering to the membrane. Nonmigrating cells on the upper surface of the membrane were removed by gentle scraping with a cotton bud and adherent migratory cells on the lower surface of the membrane were counted following fixation with 2.5% glutaraldehyde for 20 minutes and staining for 1 hour with crystal violet stain. The number of cells migrating and adhering to the underside of the membrane was evaluated by counting ten random non overlapping 40X fields. To assess the number of cells which migrated through the membrane but did not adhere to the membrane, the media in the lower chamber was collected and centrifuged at 1000 g. The cell pellet was resuspended

in 16 μL PBS, 4 μL trypan blue was added, and the cells were counted using a hemocytometer.

2.4. Visualisation of Fractionated Proteins. Protein from IEC fractions was methanol/chloroform precipitated [22] and separated using 4%–20% acrylamide gradient gels (Pierce, Rockford, Illinois) [23]. Protein gels were silver stained using a SilverSnap kit (Pierce, Rockford, Illinois).

2.5. LC-MS/MS (Liquid Chromatography-Tandem Mass Spectrometry). SDS-PAGE gel lanes were cut into 32 bands and digested in-gel with sequencing grade trypsin (Promega, Ireland) according to the method of Shevchenko et al. [15]. The resulting peptide mixtures were resuspended in 0.1% formic acid and analysed by nanoelectrospray liquid chromatography mass spectrometry (Nano-LC MS/MS). An HPLC instrument (Dionex, UK) was interfaced with an LTQ ion trap mass spectrometer (ThermoFinnigan, CA). Chromatography buffer solutions ("Buffer A", 5% acetonitrile and 0.1% formic acid; "Buffer B", 80% acetonitrile and 0.1% formic acid) were used to deliver a 60-minute gradient (35 minutes to 45% Buffer B, 10 minutes to 90%, hold for 10 minutes, 3 minutes to 5%, hold for 15 minutes). A flow rate of 2 μL/min was used at the electrospray source. Spectra were searched using the SEQUEST algorithm against the International Protein Index (IPI) database (http://www.ebi.ac.uk/IPI/IPIhelp.html), as described previously [24]. The probability-based evaluation program, Protein Prophet, was used for filtering identifications [25].

2.6. Bioinformatic Analysis. Ontological analysis of protein identifications was performed using the GOTM website (http://babelomics2.bioinfo.cipf.es/) [26]. A hypergeometric test corrected for multiple testing (Bonferroni) was used to determine statistically significant enrichments ($P < .01$). To analyse the range of proteins identified in this study, the predicted molecular weight and isoelectric point of each protein were obtained using the Compute pL/MW tool from the ExPASy server (http://expasy.org/).

2.7. Western Blot. Western blot was carried out as previously described [27] with the following modifications. Equal volumes of resting and TRAP-stimulated platelet releasates were loaded into each well and separated on a 4%–20% acrylamide gradient gel. The primary antibody for the detection of PEDF was a rabbit polyclonal antibody (1 in 1000 dilution overnight at 4°C) from Upstate Cell Signaling Solutions (Buckingham, UK). The primary antibody for the detection of involucrin was a mouse monoclonal antibody (1 in 200 dilution overnight at 4°C) from Sigma (Dublin, Ireland). Anti-mouse and anti-rabbit secondary horseradish peroxidase (HRP) conjugated antibodies were obtained from Cell Signaling Technology (Danvers, MA). Peroxidase activity was detected using luminol chemiluminescent substrate from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA).

3. Results

3.1. Strategy for Linking Biological Activities to Individual Tissue Fractions. The overall strategy is shown in Figure 1.

We used ion-exchange chromatography (IEC) to separate a complex protein mixture, proteins secreted by human platelets in response to the agonist thrombin receptor-activating peptide (TRAP), into fractions suitable for both biological analysis and identification by mass spectrometry. While strategies such as MuDPIT separate at the level of the tryptic peptide; separation at the protein level reduces sample complexity and facilitates correlation of the presence (or absence) of a protein with an observed biological function. The effect of platelet releasate on THP-1 monocyte migration was investigated to test this strategy. Addition of serum to media in the bottom of the Boyden chamber indeed stimulated migration of monocytes (data not shown), although interestingly serum stimulation did not cause the cells to adhere to the underside of the Boyden chamber membrane (Figure 2(a)). In contrast, both MCP-1 (a well-known pro-migratory chemokine for monocytes) and platelet releasate induced monocyte migration simultaneously caused these suspended cells to adhere to the underside of the membrane (Figure 2(a)). The effect of releasate was dose dependent (Figure 2(b)).

To identify releasate fractions with biological activity, consecutive IEC fractions were initially pooled in groups of three for the analysis of THP-1 migration. One pooled fraction, 25–27, was found to strongly induce monocyte migration (Figure 2(c)). We repeated the assay using the individual fractions 25, 26, and 27. In this second experiment, the pro-migratory activity could be localized to fraction 27 (Figure 2(d)). Data displayed in Figure 2 describes "adherent migration", that is, cells which migrated through the 8 μm pores and adhered to the underside of the membrane. In order to identify the proteins present in each fraction, each lane from the SDS-PAGE gels of fractionated platelet releasate (Figure 3(a)) was divided into 32 bands, digested with trypsin, and subjected to LC-MS/MS, and the corresponding proteins were identified by database search (Supplementary Table 1 in supplementary material available online at doi: 10.1155/2010/107859). For each fraction, spectral counts (counts of tandem mass spectrometry events leading to productive protein identification), an approximate indicator of protein abundance, were calculated, and the total spectral count for all proteins in the fractions was determined (Supplementary Table 2). The top fifty proteins by spectral counts are displayed in Table 1. The spectral counts (Figure 3(b)) and number of proteins identified (Figure 3(c)) correlate with each other, and also with the staining patterns of the proteins on the gels (Figure 3(a)). Of the 315 proteins identified, 220 were found in a unique fractions (70%), suggesting effective IEC separation (Figure 3(d)). Abundant proteins were often found in adjacent fractions, for example, thrombospondin was most abundant in Fraction 26 but also identified in Fractions 27 and 28. Some proteins were identified in nonadjacent fractions, perhaps reflecting differential migration of protein isoforms (e.g., splicing, proteolytic cleavage, and posttranslational modification).

3.2. Biological and Biophysical Properties of Fractionated Platelet Releasate. Our analysis resulted in the high confidence identification of 315 proteins and represents an
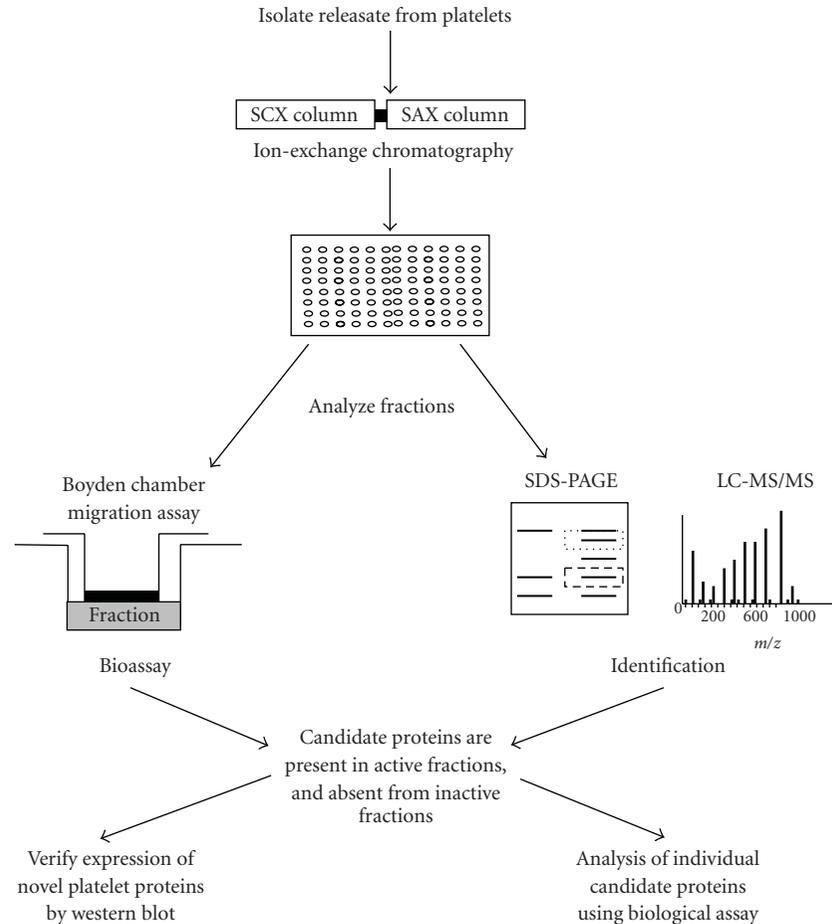
FIGURE 1: Strategy for linking function and protein identity in biofluids. An off-line MuDPIT strategy was used to identify biologically active proteins from a complex mixture. Proteins released from platelets were separated by SCX/SAX chromatography. The collected fractions were analyzed for pro-migratory activity using a monocyte Boyden chamber assay. The fractions were separated in the second dimension by SDS-PAGE, and gel lanes were excised and the proteins were digested with trypsin. The resulting peptides were separated by RP-HPLC, ionized by electrospray, and subjected to MS/MS to identify potential pro-migratory proteins.

extensive proteomic coverage of platelet releasate. The identified proteins demonstrated an extensive array of annotated biophysical properties as are evidenced by the wide range of PI and molecular weights predicted for the identifications (Figure 4(b)), suggesting that the method introduced no strong bias against any category of protein. Similarly, proteins with a wide range of annotated activities were identified. The frequency of different functional annotations within the released platelet proteins was compared to the expected frequency based on all human proteins annotated in the Gene Ontology (GO) database using the GOTM program [26]. Hemostasis, blood coagulation, and wound healing were identified as being overrepresented biological processes, as well as molecular functions such as actin binding, protein binding, and calcium ion binding (P <.01) (Figure 4(a)). GO levels between 2 and 5 were used to avoid overgeneral and overspecific functions. Many of these enriched functions appear to be relevant to potential endocrine platelet functions, for instance, "hemostasis", "wound healing", and "coagulation", while others may be relevant to the platelet activation response ("platelet activation",

"localization", "purine nucleotide binding", and "calcium binding").

3.3. Investigation of Active Fraction. Thirty-two proteins were found in Fraction 27, the only fraction that stimulated monocyte migration (Table 2). Of these proteins, six were not found in any other fraction in our study. These proteins included tissue inhibitor of metalloproteinase 1 (TIMP-1), vimentin, and pigment epithelium-derived growth factor (PEDF). Spectral counts, functions, and information on their identification in other fractions are listed for all fraction 27 proteins in Table 2. The bold areas highlight those proteins that are only found in Fraction 27. The identified proteins were compared to datasets from previous studies of platelet proteins to determine whether the identified proteins had been previously identified in platelets [1, 2, 5, 6, 27–31]. The presence of two novel platelet proteins, PEDF and involucrin (see Supplementary Table 1 for full details), was confirmed by immunoblotting (Figure 5). PEDF was identified at the expected mass of approximately 50 kDa in platelet releasate preparations with additional heavier

(a)

(b)

(c)

(d)
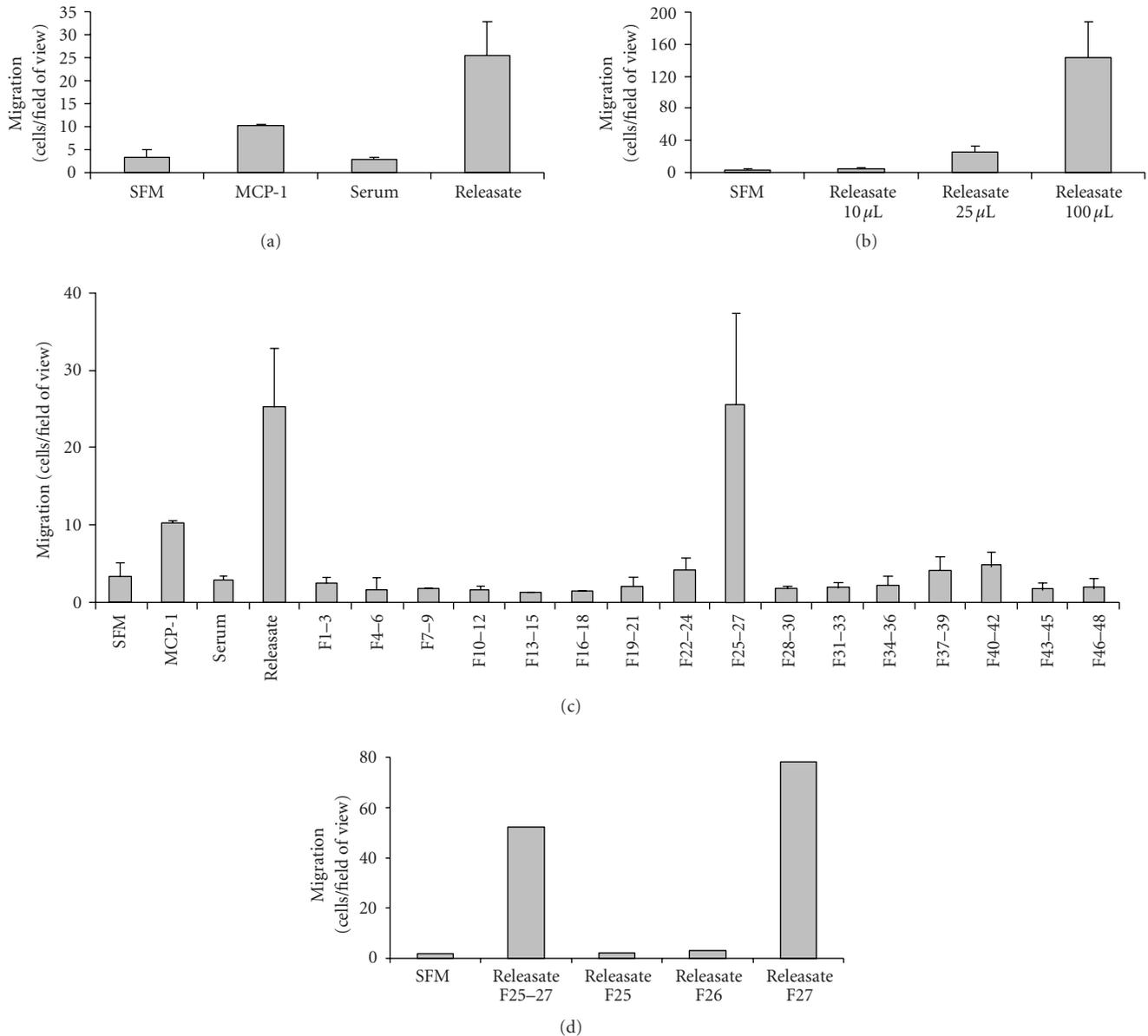
Figure 2: A pro-migratory effect identified in unfractionated and fractionated platelet releasate. THP-1 monocyte cell migration was measured using a Boyden chamber assay over 2 hours. (a) Assessment of migrated cells adhering to the underside of the membrane in response to platelet releasate, Serum (10%), and MCP-1 (100 ng/mL). $N = 3$; (b) Adherent migrated THP-1 cells in response to 10, 25 or 100 μL platelet releasate; $N = 3$. (c) Fractions were pooled in groups of three and added to serum-free media (SFM) in the bottom chamber. Positive controls include MCP-1 and unfractionated platelet releasate; $N = 3$. (d) Localisation of the pro-migratory effect of Fractions 25–27 to an individual fraction. Each fraction (2 μL) was added to serum-free media in the bottom chamber of a separate well and compared with the combination of Fractions 25–27 (2 μL of each fraction). TRAP = thrombin receptor-activating peptide. $N = 3$. The average number of migrated cells adhering to the underside of the membrane was calculated from ten 40X fields of view.

bands also detected (Figure 5(a)). Previous studies have identified PEDF at 50 kDa with additional bands at 66 kDa and 76 kDa in Y79 cell lysates using this antibody. PEDF was found to a greater extent in TRAP-stimulated platelet releasate compared to resting/unstimulated releasate. The highest band (76 kDa) was also present in the resting releasate from four different individual donors. Involucrin was identified in TRAP-stimulated platelet releasates at a mass of approximately 120 kDa (Figure 5(b)).

4. Discussion

One aim of proteomics is to address biological questions through the identification and characterization of proteins *en mass*. In this study, we developed an approach to separate a complex biological mixture using orthogonal separation methods. At the protein level, we used IEC followed by SDS-PAGE, while at the peptide level, reverse phase chromatography preceded MS analysis. We successfully applied
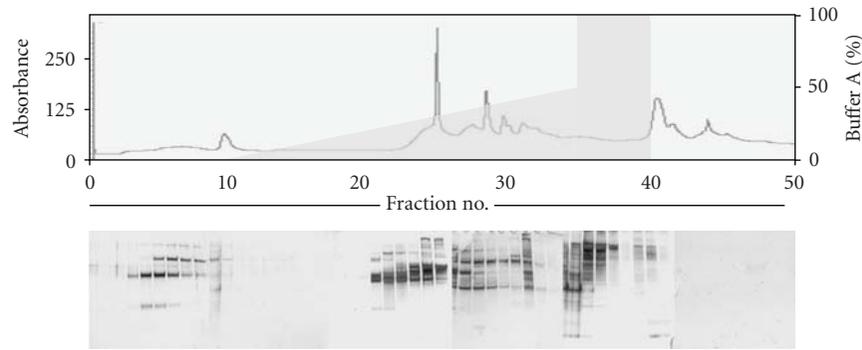
(a)
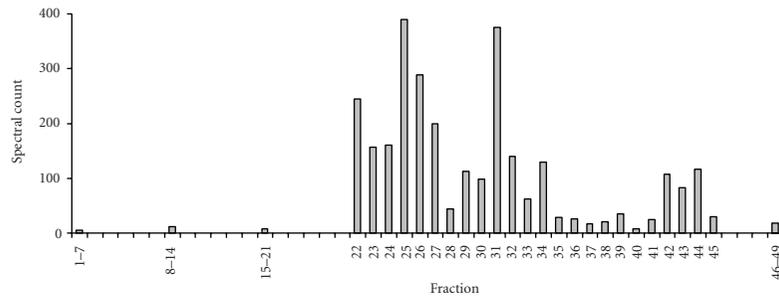
(b)
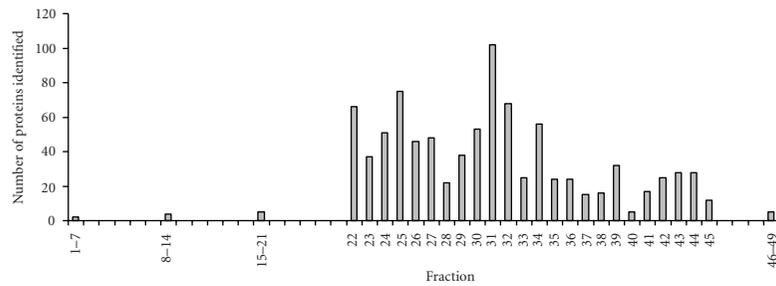
(c)
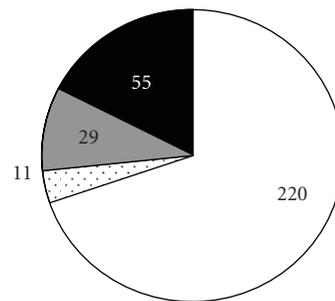
(d)

FIGURE 3: Fractionation of the platelet releasate by ion-exchange chromatography. (a) From each releasate fraction collected, 80 μL was precipitated to remove salt, resuspended in 10 μL of Laemmli buffer, and separated by MW on 4%–20% acrylamide gels. Silver-stained gels are displayed below corresponding sections of the ion-exchange UV detected chromatogram. A UV wavelength of 280 nm was used to detect eluting peptides. (b) The number of spectra detected in each fraction following LC MS/MS of tryptically digested 1D gel pieces. (c) The number of proteins identified in each fraction using SEQUEST with a score greater than 0.9. (d) Representation of the number and proportion of proteins in a single fraction, two adjacent fractions; two nonadjacent fractions, or more than two fractions.

GO biological process					Molecular function				
GO level		observed	expected	*P*-value	GO level		observed	expected	*P*-value
4	regulation of body fluids	14	1.88	5.01647E−09	3	protein binding	100	82.04	.008838235
5	hemostasis	13	1.68	1.12792E−08	3	nucleotide binding	52	33.63	.000751555
4	blood coagulation	12	1.57	4.88352E−08	4	purine nucleotide binding	48	28.78	.000228825
4	coagulation	12	1.61	6.33109E−08	5	adenyl nucleotide binding	39	23.35	.000946115
5	wound healing	12	1.75	1.67416−07	5	calcium ion binding	34	14.89	5.28281E−06
4	physiological response to wounding	19	5.88	6.93829E−06	4	cytoskeletal protein binding	21	5.88	4.74782E−07
4	response to wounding	19	6.06	1.06834E−05	5	actin binding	17	4.07	7.06132E−07
3	response to external stimulus	23	8.48	1.35735E−05	3	enzyme inhibitor activity	16	4.12	3.93645E−06
3	localization	68	43.15	4.023E−05	5	guanyl nucleotide binding	13	5.85	.006071454
4	response to unfolded protein	7	0.98	4.94329E−05	4	carrier activity	12	4.98	.004386769
4	response to protein stimulus	7	0.98	4.94329E−05	4	protease inhibitor activity	11	2.39	2.79983E−05
4	establishment of localization	67	42.74	5.55917E−05	5	endopeptidase inhibitor activity	11	2.37	2.6147E−05
4	platelet activation	4	0.31	.000205689	3	structural constituent of cytoskeleton	9	1.59	2.93143E−05
4	tissue development	12	3.69	.000335856	3	antigen binding	8	1.01	6.94074E−06
3	response to stress	28	14.03	.000359514	3	unfolded protein binding	7	1.87	.002692475
3	response to chemical stimulus	17	6.71	.000415494	4	protein binding bridging	5	1.26	.008431955
3	cell adhesion	23	11.26	.000908494	3	vitamin binding	5	1.29	.009440404
4	muscle contraction	9	2.55	.001043693	4	translation elongation factor activity	4	0.33	.000285878
3	cellular localization	22	10.83	.001255793	4	intramolecular oxidoreductase activity	4	0.72	.005532653
3	cell differentiation	21	10.27	.001492504	3	structural constituent of muscle	4	0.59	.002785384

(a)
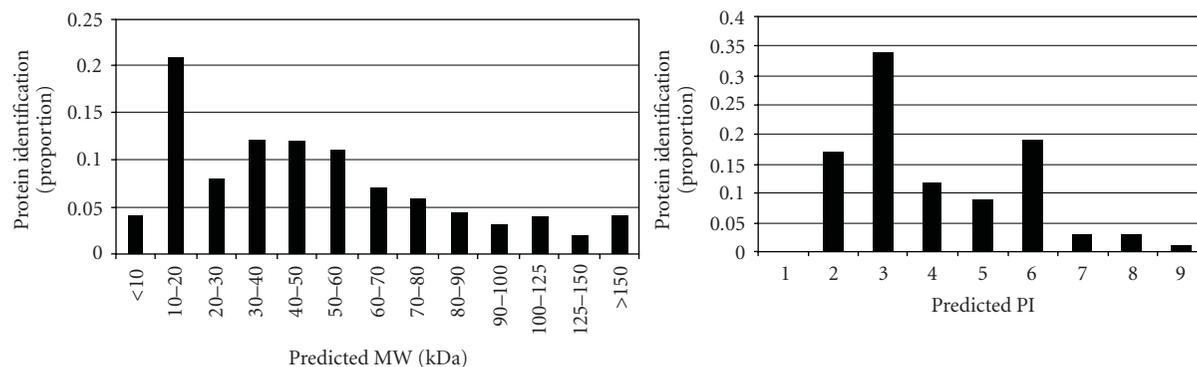
(b)

Figure 4: Bioinformatic analysis of the fractionated platelet releasate. (a) Gene ontology (GO) terms found to be significantly enriched in the experimental dataset of platelet releasate proteins ("Observed") in comparison to a reference dataset of human proteins ("Expected") along with the GO level associated with each individual ontological term. (b) The range of molecular weights (left panel) and isoelectric points (right panel) of proteins identified in this study.

protein fractions separated by IEC to a biological assay of monocyte cell migration. The denaturing conditions of gel-based protein separation largely preclude subsequent functional analysis of the proteins. However, a recently developed nondenaturing 2DE method used for the analysis of the *Shewanella oneidensis* microbial proteome may have applications for functional analysis of separated protein

mixtures in the future [32]. The enzymatic activity of malate dehydrogenases from this bacterium was retained and identified within the polyacrylamide gel following 2DE [32].

In total, 315 unique proteins were identified in this analysis of platelet releasate. Proteins which we have previously identified such as fibrinogen, albumin, thrombospondin, haemoglobin, platelet factor 4, platelet basic protein, and

Table 1: Most abundant 50 proteins determined in the study (fractions 1–49).

Accession number	Description	1-7	8-14	15-21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46-49
IPI00745872.2	Isoform 1 of serum albumin precursor	-	-	-	-	11	28	180	73	274	55	-	-	12	-	-	-	-	-	-	-	-	-	-	37	-	-	-	-
IPI00022434.4	Uncharacterized protein alb	-	-	-	28	39	82	157	32	78	21	20	22	42	29	4	11	1	4	5	1	1	3	5	38	13	-	-	-
IPI00008669.4	Uncharacterized protein ensp00000257935	-	-	-	43	75	-	15	32	-	-	26	69	7	57	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00796364.1	51 kda protein	-	-	-	-	-	-	10	55	-	62	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00021439.1	Actin; cytoplasmic 1	-	-	-	33	14	-	51	21	20	22	39	-	13	12	3	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00745872.2	Serum albumin (cell growth inhibiting protein 42)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	134	58	-
IPI00298497.3	Fibrinogen beta chain precursor	-	-	-	-	59	30	13	30	18	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	27	-	-
IPI00022434.3	Alb protein	14	11	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46	60	29
IPI00021891.5	Isoform gamma-b of fibrinogen gamma chain precursor	-	-	-	-	66	28	17	26	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00027547.2	Dermcidin precursor	-	1	2	2	7	24	6	2	15	6	10	7	31	10	-	5	-	-	-	-	-	-	-	11	-	-	2	-
IPI00884222.1	Similar to keratin, type 1 cytoskeletal 10	-	-	-	-	-	-	-	62	-	-	-	54	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00021841.1	Apolipoprotein a-i precursor	-	-	-	-	-	-	15	-	3	-	-	1	4	-	2	-	-	-	-	-	-	-	-	-	-	67	9	-
IPI00013933.2	Isoform dpi of desmoplakin	-	-	-	27	-	4	-	-	5	-	2	6	32	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00784154.1	60 kda heat shock protein; mitochondrial precursor	-	-	-	60	2	-	-	-	-	-	-	-	2	4	14	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00296099.6	Thrombospondin-1 precursor	-	-	-	-	-	-	-	-	37	32	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00016801.1	Glutamate dehydrogenase 1; mitochondrial precursor	-	-	-	28	28	-	-	-	-	-	-	-	16	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00021885.1	Isoform 1 of fibrinogen alpha chain precursor	-	-	-	-	19	12	17	19	5	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00478003.1	Alpha-2-macroglobulin precursor	-	-	-	-	-	-	3	-	9	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46	-	-
IPI00022445.1	Platelet basic protein precursor	-	-	-	-	2	5	3	11	12	2	2	8	3	4	3	2	-	-	-	-	-	-	-	-	-	-	-	-
IPI00419424.3	Igkv1-5 protein	-	-	-	-	1	2	7	-	2	-	-	-	3	-	-	-	-	-	-	-	-	-	-	2	-	33	1	-
IPI00021885.1	Splice isoform 1 of fibrinogen alpha chain precursor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	42	9	-
IPI00217960.1	Isoform 2 of camp-dependent PK; alpha-catalytic subunit	-	-	-	17	12	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00019502.3	Myosin-9	-	-	-	-	-	-	-	19	-	-	1	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00164623.4	Complement c3 precursor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	37	-	-
IPI00022446.1	Platelet factor 4 precursor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	6	-	7	22	-
IPI00022463.1	Serotransferrin precursor	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	3	-	30	-	-
IPI00453473.6	Histone h4	-	-	-	-	2	4	-	4	3	3	-	5	6	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00021536.1	Calmodulin-like protein 5	-	-	-	-	-	5	12	-	-	-	-	7	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00298994.6	Uncharacterized protein tln1	-	-	-	-	-	-	-	-	31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00219221.3	Galectin-7	-	-	-	8	3	2	6	-	-	-	-	-	8	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00384938.1	Hypothetical protein dkfzp686n02209	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	26	-	-
IPI00783987.2	Complement c3 precursor (fragment)	-	-	-	4	-	-	-	6	-	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00015614.3	Isoform a of trypsin-3 precursor	-	-	-	-	-	-	-	-	-	-	-	-	2	-	3	5	7	-	4	-	3	-	-	-	-	-	-	-
IPI00431645.1	Hp protein	-	-	-	-	5	9	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00302592.2	Filamin a; alpha isoform 1	-	-	-	-	-	-	-	23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00004358.4	Glycogen phosphorylase; brain form	-	-	-	19	-	-	-	-	-	-	-	-	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00022204.2	Serpin b3	-	-	-	-	-	5	9	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00027462.1	Protein s100-a9	-	-	-	-	-	4	16	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00021891.5	Splice isoform gamma-b of fibrinogen gamma chain precursor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21	-	-
IPI00154742.6	Iglc1 protein	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-
IPI00291175.7	Isoform 1 of vinculin	-	-	-	-	-	-	-	20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00553177.1	Alpha-1-antitrypsin precursor	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	-	-
IPI00025753.1	Desmoglein-1 precursor	-	-	-	5	-	3	-	-	-	-	2	6	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00026314.1	Isoform 1 of gelsolin precursor	-	-	-	-	-	-	-	-	3	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00017601.1	Ceruloplasmin precursor	-	-	-	1	-	-	-	-	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00003935.6	Histone h2b type 2-e	-	-	-	5	2	-	-	-	1	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00383164.1	Snc66 protein	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	-	-
IPI00008603.1	Actin; aortic smooth muscle	-	3	2	-	-	-	-	-	-	-	2	-	4	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00010951.2	Epiplakin	-	-	-	5	-	-	2	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IPI00654755.3	Hemoglobin subunit beta	-	-	-	-	1	4	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

(a)

(b)
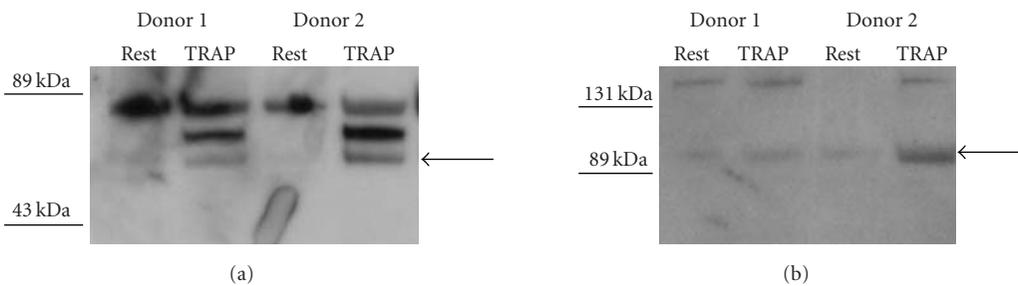
Figure 5: Western blot validation of novel platelet proteins. 30 μL of resting and TRAP-stimulated platelet releasate were separated on a 4%–20% gradient SDS-PAGE gel and probed for PEDF and involucrin. (a) PEDF was detected as a band at approximately 50 kDa, with additional heavier bands that may represent glycosylated forms of the protein. (b) Involucrin was detected at the expected molecular weight as a band resolving at approximately 120 kDa. The locations of molecular weight markers are also indicated. These proteins were identified in TRAP-stimulated platelet releasates from 4 individual donors. Images displayed are representative of the 4 individual donors. R = Resting/unstimulated platelet releasate T = TRAP-stimulated platelet releasate.

TABLE 2: Proteins identified in Fraction 27.

IPI no.	Protein name	Spectral count	Other fractions where protein was found	Description	Previously found in platelets
IPI00745872.2	Isoform 1 of serum albumin precursor	274	23–26, 28, 31, 42	Serum albumin	Yes
IPI00022434.4	Uncharacterized protein alb	78	22–26, 28–43,	Albumin protein	Yes
IPI00021439.1	Actin, cytoplasmic 1	20	22-23, 25-26, 28–29, 31–33	Highly conserved cytoskeletal protein involved in cell motility	Yes
IPI00298497.3	Fibrinogen beta chain, precursor	18	23–26, 28, 44	Yields monomers that polymerize fibrin and act as a cofactor in platelet aggregation	Yes
IPI00021891.5	Isoform gamma-b of fibrinogen gamma chain precursor	14	23–26	Yields monomers that polymerize fibrin and act as a cofactor in platelet aggregation	Yes
IPI00027547.2	Dermcidin precursor	15	8–26, 28–32, 34, 41, 45	Antimicrobial activity and survival promoting peptide for neurons	
IPI00884222.1	Similar to keratin, type 1 cytoskeletal 10	62	32	Heterotetramer of two type I and two type II keratins	
IPI00021841.1	Apolipoprotein a-i precursor	3	25, 31-32, 34, 44–45	Participates in reverse transport of cholesterol from tissues to the liver	
IPI00013933.2	Isoform dpi of desmoplakin	5	22, 24, 29–32	High molecular weight protein of desmosomes	
IPI00296099.6	Thrombospondin-1 precursor	32	26, 28	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions	Yes
IPI00021885.1	Isoform 1 of fibrinogen alpha chain precursor	5	23–26, 34	Cofactor in platelet aggregation and yield monomers that polymerize into fibrin	Yes
IPI00478003.1	Alpha-2-macroglobulin precursor	13	23, 25, 44	Protease inhibitor	
IPI00022445.1	Platelet basic protein precursor	12	23–26, 28–34	Precursor for a number of peptides with antimicrobial (TC-1 and TC-2), chemoattractant, and stimulation of DNA synthesis activity	Yes
IPI00419424.3	Igkv1-5 protein	2	23–25, 31, 42, 44–45	Immunoglobulin	
IPI00453473.6	Histone h4	3	22-23, 25-26, 30–32, 34, 36	Core component of nucleosomes playing a central role in DNA repair, replication, and transcription	
IPI00783987.2	Complement c3 precursor (fragment)	15	22, 25	Involved in activation of the complement system and local inflammation	
IPI00026314.1	Isoform 1 of gelsolin precursor	14	26	Calcium-regulated, actin-modulating protein	Yes
IPI00022434.4	Uncharacterized protein alb.	78	22–26, 28–43	Uncharacterized albumin protein	Yes

TABLE 2: Continued.

IPI no.	Protein name	Spectral count	Other fractions where protein was found	Description	Previously found in platelets
IPI00017601.1	Ceruloplasmin precursor	15	22	Glycoprotein involved in iron transport across cell membrane	
IPI00382606.1	Factor vii active site mutant immunoconjugate	4	45	Initiates extrinsic pathway of blood coagulation	
IPI00154742.6	Igl@ protein	2	24-25, 31	Immunoglobulin	
IPI00550991.3	Alpha-1-antichymotrypsin precursor	6	44	Inhibits neutrophil cathepsin G and mast cell chymase	
IPI00025252.1	Protein disulfide-isomerase a3 precursor	4	26	Catalyses the rearrangement of S-S bonds in proteins	Yes
IPI00022431.1	Alpha-2-hs-glycoprotein precursor	3	44	Secreted protein, part of the fetuin family, may be involved in differentiation	Yes
IPI00032292.1	Metalloproteinase inhibitor 1 precursor	3		Complexes with metalloproteinases and irreversibly inactivates them	Yes
IPI00555812.4	Vitamin d-binding protein precursor	1	23, 44	Multifunctional protein found in many body fluids and on the surface of various cell types	Yes
IPI00179709.4	Isoform 1 of tubulin alpha-3c/d chain	2		Major constituent of microtubules	
IPI00006114.4	Proliferation-inducing protein 35/PEDF	2		Induces neuronal differentiation and inhibits angiogenesis	Yes
IPI00739464.1	Similar to cytoplasmic beta-actin	1	31	Highly conserved cytoskeletal protein involved in cell motility	Yes
IPI00414860.6	60s ribosomal protein l37a	1		Ribosomal protein containing a zinc finger domain, expressed in many cell types	Yes
IPI00020996.3	Insulin-like growth factor-binding protein complex acid labile chain precursor	1		Secreted protein involved in protein complexes with insulin-like growth factor family	
IPI00418471.6	Vimentin	1		Class III intermediate filament protein found in many nonepithelial cells	Yes

actin binding proteins were well represented [1]. While it cannot be discounted that some of these proteins may be residual serum proteins or proteins derived from cellular debris, the identifications detailed in this study are in agreement with previous publications from our laboratory and further afield (see [1, 2, 4, 27, 33–36]).

We also report the identification of proteins which have previously been implicated in neuronal exocytosis such as calmodulin-like protein 3 in Fraction 25 [37] and pleckstrin and sec7 domain protein in Fractions 22 and 26, respectively, [38]. Pigment epithelium-derived factor (PEDF) was also identified in the platelet releasate by mass spectrometric methods and its presence was validated by western blot. PEDF is a potent inhibitor of angiogenesis in the eye [39] and has also been detected in plasma and serum [40, 41]. Interestingly, a recent study suggests that pro- and anti-angiogenic proteins within the platelet may be separated into distinct subsets of alpha granules which are differentially released depending on the agonist [42]. We also confirmed the presence of involucrin, a protein found as a component of scaffolding in the cell envelope of stratified epithelium, in the platelet releasate by western blotting. This protein has not

been previously identified in nonepithelial cell types to the best of our knowledge.

Previous studies in our laboratory established that the platelet releasate had a much greater pro-migratory effect on THP-1 monocytes than on human arterial smooth muscle cells or human retinal endothelial cells (data not shown). The use of ion-exchange fractionation enabled us to locate the pro-migratory activity of the platelet releasate to just one of 49 Fractions. Candidate proteins which may be responsible for the monocyte pro-migratory activity of Fraction 27 include tissue inhibitor of metalloproteinase 1 (TIMP-1) and proliferation-inducing protein 35 (PEDF). Future studies will focus on identifying the potential pro-migratory protein, or proteins, in Fraction 27 by using neutralizing antibodies in our cell-based assay, against some of the identified proteins. A decrease in the biological activity of the candidate protein due to the presence of a neutralizing antibody would further support its pro-migratory effect in our cell assay.

However it is also possible, despite the extensive protein separation methods used, that the protein responsible for the migratory activity is a small protein of low abundance which may never be captured by mass spectrometry. Antibody arrays of growth factors and cytokines may also assist in the identification of the pro-migratory protein secreted by platelets. Using this technology we have previously shown that PDGF, angiogenin, RANTES, and many other cytokines and growth factors are secreted by platelets following thrombin activation [27]. These proteins were not identified in any fraction in the current mass spectrometric study of the platelet releasate. Investigation of the contribution of some likely pro-migratory cytokine candidates within the platelet releasate has ruled out certain proteins in terms of monocyte migration. Anti-RANTES neutralizing antibodies had no effect on platelet releasate-induced THP-1 monocyte migration (data not shown). It seems unlikely that MCP-1 is the chemokine responsible for platelet releasate-induced THP-1 migration because 100 ng/mL MCP-1 induced a much lower level of THP-1 migration than TRAP-stimulated platelet releasate.

Both collagen I and collagen IV were identified in Fraction 27 but not in any other fraction, and previous studies in our laboratory have shown a pro-migratory effect of collagen I on arterial smooth muscle cells, so we also investigated the effect of these extracellular matrix proteins on THP-1 migration. However, no evidence of a pro-migratory effect was observed (data not shown). In addition, the possibility that a group of proteins rather than an individual protein in Fraction 27 contributed to the pro-migratory effect can not be discounted without further investigation.

In summary, the approach outlined in the current study facilitated the penetration of the platelet releasate proteome to a sufficient degree to increase the number of low-abundance identifications. This in turn highlighted candidate proteins responsible for platelet releasate induced monocyte cell migration which merit further targeted functional analysis. The method described here could be equally applied to other secretory systems and biological situations involving the functional analysis of secreted proteins from one cell type on another cell function.

References

[1] J. A. Coppinger, G. Cagney, S. Toomey, et al., "Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions," *Blood*, vol. 103, no. 6, pp. 2096–2104, 2004.

[2] A. Garcia, S. Prabhakar, C. J. Brock, et al., "Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry," *Proteomics*, vol. 4, no. 3, pp. 656–668, 2004.

[3] K. Gevaert, M. Goethals, L. Martens, et al., "Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides," *Nature Biotechnology*, vol. 21, pp. 566–569, 2003.

[4] P. Gravel, J. C. Sanchez, C. Walzer, et al., "Human blood platelet protein map established by two-dimensional polyacrylamide gel electrophoresis," *Electrophoresis*, vol. 16, no. 7, pp. 1152–1159, 1995.

[5] J. Moebiust, R. P. Zahedi, U. Lewandrowski, C. Berger, U. Walter, and A. Sickmann, "The human platelet membrane proteome reveals several new potential membrane proteins," *Molecular and Cellular Proteomics*, vol. 4, no. 11, pp. 1754–1761, 2005.

[6] L. Guerrier, S. Claverol, F. Fortis, et al., "Exploring the platelet proteome via combinatorial, hexapeptide ligand libraries," *Journal of Proteome Research*, vol. 6, no. 11, pp. 4290–4303, 2007.

[7] L. Liu, D. Hartwig, S. Harloff, et al., "Corneal epitheliotrophic capacity of three different blood-derived preparations," *Investigative Ophthalmology and Visual Science*, vol. 47, no. 6, pp. 2438–2444, 2006.

[8] J. S. Rhee, M. Black, U. Schubert, et al., "The functional role of blood platelet components in angiogenesis," *Thrombosis and Haemostasis*, vol. 92, no. 2, pp. 394–402, 2004.

[9] A. Brill, H. Elinav, and D. Varon, "Differential role of platelet granular mediators in angiogenesis," *Cardiovascular Research*, vol. 63, no. 2, pp. 226–235, 2004.

[10] L. Anderson and N. G. Anderson, "High resolution two-dimensional electrophoresis of human plasma proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5421–5425, 1977.

[11] A. J. Link, J. Eng, D. M. Schieltz, et al., "Direct analysis of protein complexes using mass spectrometry," *Nature Biotechnology*, vol. 17, no. 7, pp. 676–682, 1999.

[12] J. M. Asara, X. Zhang, B. Zheng, et al., "In-gel stable isotope labeling for relative quantification using mass spectrometry," *Nature Protocols*, vol. 1, no. 1, pp. 46–51, 2006.

[13] J. E. Froehlich, C. G. Wilkerson, W. K. Ray, et al., "Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis," *Journal of Proteome Research*, vol. 2, no. 4, pp. 413–425, 2003.

[14] K. Kaneko, H. Ando, T. Seo, Y. Ono, T. Tainaka, and W. Sumida, "Proteomic analysis of protein plugs: causative agent

of symptoms in patients with choledochal cyst," *Digestive Diseases and Sciences*, vol. 52, no. 8, pp. 1979–1986, 2007.

[15] A. Shevchenko, M. Wilm, O. Vorm, and M. Mann, "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels," *Analytical Chemistry*, vol. 68, no. 5, pp. 850–858, 1996.

[16] A. Leitner and W. Lindner, "Chemistry meets proteomics: the use of chemical tagging reactions for MS-based proteomics," *Proteomics*, vol. 6, no. 20, pp. 5418–5434, 2006.

[17] C. M. Salisbury and B. F. Cravatt, "Activity-based probes for proteomic profiling of histone deacetylase complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 4, pp. 1171–1176, 2007.

[18] G. S. Magalhaes, M. Lopes-Ferreira, I. L. M. Junqueira-de-Azevedo, et al., "Natterins, a new class of proteins with kininogenase activity characterized from Thalassophryne nattereri fish venom," *Biochimie*, vol. 87, no. 8, pp. 687–699, 2005.

[19] H. Wang, X. Yang, G. C. Bowick, et al., "Identification of proteins bound to a thioaptamer probe on a proteomics array," *Biochemical and Biophysical Research Communications*, vol. 347, no. 3, pp. 586–593, 2006.

[20] J. Biarc, I. S. Nguyen, A. Pini, et al., "Carcinogenic properties of proteins with pro-inflammatory activity from Streptococcus infantarius (formerly S. bovis)," *Carcinogenesis*, vol. 25, no. 8, pp. 1477–1484, 2004.

[21] M. E. Mendelsohn, Y. Zhu, and S. O'Neill, "The 29-kDa proteins phosphorylated in thrombin-activated human platelets are forms of the estrogen receptor-related 27-kDa heat shock protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 24, pp. 11212–11216, 1991.

[22] D. Wessel and U. I. Flugge, "A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids," *Analytical Biochemistry*, vol. 138, no. 1, pp. 141–143, 1984.

[23] U. K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–685, 1970.

[24] J. R. Yates III, J. K. Eng, A. L. McCormack, and D. Schieltz, "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database," *Analytical Chemistry*, vol. 67, no. 8, pp. 1426–1436, 1995.

[25] A. I. Nesvizhskii, A. Keller, E. Kolker, and R. Aebersold, "A statistical model for identifying proteins by tandem mass spectrometry," *Analytical Chemistry*, vol. 75, no. 17, pp. 4646–4658, 2003.

[26] B. Zhang, D. Schmoyer, S. Kirov, and J. Snoddy, "GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using gene ontology hierarchies," *BMC Bioinformatics*, vol. 5, article 16, 2004.

[27] J. A. Coppinger, R. O'Connor, K. Wynne, et al., "Moderation of the platelet releasate response by aspirin," *Blood*, vol. 109, no. 11, pp. 4786–4792, 2007.

[28] B. Walkowiak, M. Kaminska, W. Okroj, et al., "The blood platelet proteome is changed in UREMIC patients," *Platelets*, vol. 18, no. 5, pp. 386–388, 2007.

[29] D. M. Maynard, H. F. Heijnen, M. K. Horne, J. G. White, and W. A. Gahl, "Proteomic analysis of platelet α-granules using mass spectrometry," *Journal of Thrombosis and Haemostasis*, vol. 5, no. 9, pp. 1945–1955, 2007.

[30] L. Martens, P. Van Damme, J. Van Damme, et al., "The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile," *Proteomics*, vol. 5, pp. 3193–3204, 2005.

[31] B. A. Garcia, D. M. Smalley, H. Cho, J. Shabanowitz, K. Ley, and D. F. Hunt, "The platelet microparticle proteome," *Journal of Proteome Research*, vol. 4, no. 5, pp. 1516–1521, 2005.

[32] C. S. Giometti, T. Khare, S. L. Tollaksen, et al., "Analysis of the Shewanella oneidensis proteome by two-dimensional gel electrophoresis under nondenaturing conditions," *Proteomics*, vol. 3, pp. 777–785, 2003.

[33] V. Dubernard, B. B. Arbeille, M. B. Lemesle, and C. Legrand, "Evidence for an α-granular pool of the cytoskeletal protein α-actinin in human platelets that redistributes with the adhesive glycoprotein thrombospondin-1 during the exocytotic process," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 17, no. 10, pp. 2293–2305, 1997.

[34] D. V. Gnatenko, J. J. Dunn, S. R. McCorkle, D. Weissmann, P. L. Perrotta, and W. F. Bahou, "Transcript profiling of human platelets using microarray and serial analysis of gene expression," *Blood*, vol. 101, no. 6, pp. 2285–2293, 2003.

[35] G. O. Gogstad, I. Hagen, R. Korsmo, and N. O. Solum, "Characterization of the proteins of isolated human platelet alpha-granules. Evidence for a separate alpha-granule-pool of the glycoproteins IIb and IIIa," *Biochim Biophys Acta*, vol. 670, no. 2, pp. 150–162, 1981.

[36] J. P. McRedmond, S. D. Park, D. F. Reilly, J. A. Coppinger, P. B. Maguire, D. C. Shields, and D. J. Fitzgerald, "Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes," *Molecular and Cellular Proteomics*, vol. 3, no. 2, pp. 133–144, 2004.

[37] S. Quetglas, C. Iborra, N. Sasakawa, et al., "Calmodulin and lipid binding to synaptobrevin regulates calcium-dependent exocytosis," *EMBO Journal*, vol. 21, no. 15, pp. 3970–3979, 2002.

[38] J. Polgar, S. H. Chung, and G. L. Reed, "Vesicle-associated membrane protein 3 (VAMP-3) and VAMP-8 are present in human platelets and are required for granule secretion," *Blood*, vol. 100, no. 3, pp. 1081–1083, 2002.

[39] D. W. Dawson, O. V. Volpert, P. Gillis, et al., "Pigment epithelium-derived factor: a potent inhibitor of angiogenesis," *Science*, vol. 285, no. 5425, pp. 245–248, 1999.

[40] S. V. Petersen, Z. Valnickova, and J. J. Enghild, "Pigment-epithelium-derived factor (PEDF) occurs at a physiologically relevant concentration in human blood: purification and characterization," *Biochemical Journal*, vol. 374, no. 1, pp. 199–206, 2003.

[41] A. J. Jenkins, S. X. Zhang, K. G. Rowley, et al., "Increased serum pigment epithelium-derived factor is associated with microvascular complications, vascular stiffness and inflammation in type 1 diabetes1," *Diabetic Medicine*, vol. 24, no. 12, pp. 1345–1351, 2007.

[42] J. E. Italiano Jr., J. L. Richardson, S. Patel-Hett, et al., "Angiogenesis is regulated by a novel mechanismml: pro- and anti-angiogenic proteins are organized into separate platelet {α}-granules and differentialy released," *Blood*, 2007.

Research Article

Statistical Analysis of Variation in the Human Plasma Proteome

Todd H. Corzett,[1] Imola K. Fodor,[2] Megan W. Choi,[1] Vicki L. Walsworth,[1] Kenneth W. Turteltaub,[1] Sandra L. McCutchen-Maloney,[1] and Brett A. Chromy[1]

[1] *Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA*

[2] *Department of Biostatistics, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*

Correspondence should be addressed to Brett A. Chromy, chromy1@llnl.gov

Quantifying the variation in the human plasma proteome is an essential prerequisite for disease-specific biomarker detection. We report here on the longitudinal and individual variation in human plasma characterized by two-dimensional difference gel electrophoresis (2-D DIGE) using plasma samples from eleven healthy subjects collected three times over a two week period. Fixed-effects modeling was used to remove dye and gel variability. Mixed-effects modeling was then used to quantitate the sources of proteomic variation. The subject-to-subject variation represented the largest variance component, while the time-within-subject variation was comparable to the experimental variation found in a previous technical variability study where one human plasma sample was processed eight times in parallel and each was then analyzed by 2-D DIGE in triplicate. Here, 21 protein spots had larger than 50% CV, suggesting that these proteins may not be appropriate as biomarkers and should be carefully scrutinized in future studies. Seventy-eight protein spots showing differential protein levels between different individuals or individual collections were identified by mass spectrometry and further characterized using hierarchical clustering. The results present a first step toward understanding the complexity of longitudinal and individual variation in the human plasma proteome, and provide a baseline for improved biomarker discovery.

1. Introduction

Mapping the human proteome presents a significant scientific challenge, partly because of the complexity of the population and partly because of technological limitations [1]. However, potential rewards in the diagnosis and treatment of diseases make proteomic characterization of human plasma a very worthwhile endeavor. The Human Proteome Organization (HUPO) represents an international consortium of academic and industrial partners whose common goal is to foster collaboration and facilitate a better understanding of the human proteome. Recognizing the need for reproducibility, and following in the footsteps of the more mature field of gene expression analysis, proteomic standards are starting to emerge [2, 3]. The Human Plasma Proteome (HPP) project [4] of HUPO, which specifically targets plasma proteins, has made considerable progress while highlighting the complexity of plasma proteomics. For example, protein identification of the same specimen resulted in less than 50% agreement when repeated multiple times [5, 6], reflecting the challenges involved in biomarker discovery from human plasma [7, 8] and underlining the need for improvements in plasma proteomic characterization. Studies providing prefractionation and other sample preparation aspects are looking to improve this process [9–12].

A primary technological problem that needs to be addressed is the quantification of the experimental variation on a given proteomic platform. Next, the baseline variation within individuals over time and the variation between multiple individuals also need to be quantitated. Searching for disease-specific biomarkers makes sense only after these two steps are addressed. Our recent study, referred to as the Technical Variation Study (TVS) [13] throughout the manuscript, addressed the first question for two-dimensional difference gel electrophoresis (2-D DIGE) experiments by processing one human plasma sample eight

times and analyzing each of the resulting eight technical replicates in triplicate on twelve gels [13–16]. The present study is a follow-up to the TVS, whereby plasma samples from eleven healthy volunteer subjects, taken at three time points separated by two weeks, were analyzed in triplicate on 50 gels.

The goal of this study was to assess longitudinal and individual variation in human plasma and to compare results to the experimental variation detected in the previously reported TVS [13]. While differences were detected in the plasma proteome within individuals over time, our analyses indicate that individual variation contributes the largest observed proteomic variability. Further, our results demonstrate that gender-related proteomic differences can be detected by 2-D DIGE and should also be considered in biomarker discovery. Overall, this work represents a first step in quantitating the variability in human plasma by addressing the individual and longitudinal proteomic variation in human plasma.

2. Materials and Methods

2.1. Sample Collection. Blood samples were collected from eleven healthy volunteers (five males, six females) at three time points separated by two weeks, with informed consent under Institutional Review Board approval from Lawrence Livermore National Laboratory. To minimize the effect of daily variations within an individual, the samples from a given subject were taken at approximately the same time, within a thirty-minute window, in the morning for each time point. Other variables were not controlled for in order to better mimic the variability in typical human plasma samples (age, fasting, illness, medication, etc.). To better examine the longitudinal variation and minimize the chance of an individual providing samples while experiencing an underlying condition such as a cold, two weeks between sample collection were chosen. Each individual was assigned an identification number to blind the samples and ease the experimental design.

2.2. Top-6 High-Abundance Protein Depletion and Sample Preparation. To increase the resolution of the 2-D DIGE technique, the six most abundant plasma proteins were depleted using affinity chromatography, as previously reported [12]. The sample cleanup and protein assay was performed as described previously [12, 13, 17, 18].

2.3. 2-D DIGE and Gel Imaging. The 33 top-6-depleted plasma samples from the 11 individuals were analyzed in triplicate in a 50-gel 2-D DIGE experiment [13–16] (see Table 1 in Supplementary Material available online at doi: 10.1155/2010/258494). Each gel contained three samples, one internal pooled standard and two experimental samples. The internal pooled standard consists of an equal amount of each of the 33 samples and was labeled with the Cy2 dye (GE Healthcare). Each experimental sample was dye-swapped and labeled with both the Cy3 and the Cy5 dyes (GE Healthcare) in the experimental design to mitigate the effect

of potential dye-specific variations [19]. Samples from individuals obtained at different times were compared on some gels, while samples from two different individuals were compared on other gels. Gels were run randomly in batches of twelve in order to minimize batch-to-batch variability. Supplementary Table 1 shows the complete experimental design. Labeling first dimension (pI) separation, second dimension (mw) separation, and gel imaging was performed as described previously [13]. Mass spectrometry was carried out as previously described [20].

2.4. Data Analysis. The DeCyder Differential Analysis Software v5.01 (GE Healthcare) was used for quantitating differential abundance of proteins. The Differential In-gel Analysis (DIA) module was used to determine the optimal spot detection settings. Images were loaded into the Batch Processor module with the estimated number of spots set to 2,500. The master gel was assigned automatically to the gel with the most spots detected. Each sample was grouped for analysis in the Biological Variation Analysis (BVA) module. During batch processing, the Cy2 channel from each gel was used for normalization of the spot intensities and for automated matching between gels. For each spot on each gel, the software reported the standardized abundance (SA) as the ratio of the volume in the Cy3 (or Cy5) sample to the volume of the pooled standard sample labeled with Cy2, where the volumes were normalized across the gels. Standardized log abundance (SLA), defined as $\log_{10}$(SA), was used in quantifying differential expression. Fold change between groups was calculated as the ratio of the average SA in the two groups. If R denotes that ratio, the fold change F was defined as $F = R$ if $R \geq 1$ and $F = -1/R$ otherwise. A k-fold expression increase/decrease corresponded to a $+k/-k$ value of F.

Using DeCyder, all possible pairwise comparisons were made between the 33 groups defined by the eleven subjects at the three times. Within the BVA module each comparison was filtered to find the spots having (a) P-value $\leq .05$ and (b) greater than 1.5-fold change in expression between the groups. The analysis was converted into DeCyder 2D (v6.5), and the Extended Data Analysis (EDA) module (GE Healthcare) was used to perform expression pattern clustering [21–26]. Data from the TVS were integrated into the analysis, pooled standards were normalized and principal component analysis was conducted [27–30] on the spots that were successfully matched on >75% of the gels from the TVS and the present study [13].

Spot characteristics calculated by DeCyder were exported for further statistical processing into the R statistical computing environment [19] (http://www.r-project.org/). Summary statistics for the spot matching across the gels were calculated. The high-quality spots, defined as the spots matched on at least 75% of the gels, were subjected to further analysis. Spotwise standard deviations (SDs) of the SLA values and coefficients of variations (CVs) of the SA values were calculated, first by using all the data at a given spot to obtain one SD and one CV for that spot, then by performing the same calculations separately for the eleven subjects, thus

obtaining eleven SD and CV values for each spot. The former method estimated the protein expression variation among all subjects and time points, while the latter addressed the variation within individuals through time. The results were compared to the spotwise SD and CV values obtained from the TVS [13].

To quantitate the relative contribution of the components of variation, mixed-effects statistical modeling [31] was performed. Let y_{ijk} denote the SLA at spot i on gel j measured with dye k, with $i = 1, \ldots, I$, where I represents the number of spots matched on 75% of the gels, $j = 1, \ldots, 50$, and $k = 1, 2$. In addition, let $l = 1, \ldots, 11$ and $m = 1, 2, 3$ indicate the subject and time indices, respectively. Let $g = 1, 2$ indicate male and female genders, respectively. The assumed model was of the form

$$y_{ijklm} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \gamma_{ig} + a_{il} + b_{il(m)} + e_{ijklm}, \quad (1)$$

where μ is the overall mean, α_j denote the coefficients for the fixed gel effects, β_k the coefficients for the fixed dye effects, $(\alpha\beta)_{jk}$ the coefficients for the gel-dye interactions, γ_{ig} the coefficient for the fixed gender effect at spot i, and a_{il}, $b_{il(m)}$, and e_{ijklm} the random effects components for subject (individual), time (longitudinal), and error at spot i, independent and normally distributed [32] with mean zero and variance σ_{si}^2, σ_{ti}^2, and σ_{ei}^2, respectively. The gender, subject, and time subscripts in (1) are redundant, as for any given j and k, the identity of the sample, including the subject, gender and time, was known. However, we included them for the clarity of the model description. Since the gel and dye factors were balanced with respect to the spots (the first four terms in the model were common to all spots), (1) was fit in two stages, with results equivalent to, and computationally more efficient than, the full one-stage solution in (1) repeated at each spot. Similar methods have been established for microarrays [33]. In the first stage, the data from all I spots were used to estimate the global dye and gel effects; that is, only the first four terms in (1) plus error were included in the model. In the second stage, the last four terms in (1) were fit to the residuals from the first stage, one spot at a time. In essence, this first stage amounted to a normalization step, whereby the fixed dye and gel effects were estimated and removed by pooling the information across all the spots. In the second stage, a fixed gender effect and random variance components of subject, time, and error were estimated separately at each spot. Thus, at spot i, the total variance σ_{yi}^2 was separated into its random components as $\sigma_{yi}^2 = \sigma_{si}^2 + \sigma_{ti}^2 + \sigma_{ei}^2$.

The effect of additional statistical normalizations of the SLA on the variance component estimates and on the differentially expressed spots was investigated. The SLA values obtained from DeCyder were further normalized by statistical methods that corrected for potential dye biases within gels and range differences among the gels as previously described [34].

The spots that were determined to be of differential abundance (>1.5-fold difference with P-value <.05) were excised from the pick gel and identified by mass spectrometry as previously reported [12, 18, 20]. Identified spots were selected in DeCyder for additional expression pattern clustering.

3. Results and Discussion

3.1. Experimental Design.
The experimental design is shown in Supplementary Table 1 and Supplementary Figure 1. Rather than randomly pairing the samples on the gels, the design was selected to minimize the experimental variation among the samples whose comparison was of most interest. By placing the samples from a subject across different time points on the same gel, gel-related variations for intrasubject comparisons were minimized. Essentially, our design was based on the requirements that (1) each of the 33 samples has three replicates and (2) comparisons of the samples from the same subject were of more interest than comparisons of different subjects across time points. This led to an experimental design that contained 22 gels used for comparing the same individual at different time points and 28 gels to compare two individuals at different time points. In addition, the use of dye swapping and triplicates also contributed to overall quality of the data. Our results suggest that gender variability may also be present. As individual and longitudinal variability was our main objective, we did not attempt to control for gender differences. We selected both males and females for our study to get a more appropriate human sample set. Future work looking at human proteome variability should account for gender-specific variability in the design of the experiment.

3.2. Spot Matching.
Landmarks were placed manually on each gel to assist in the spot matching across the gels. Spots of interest identified through the analyses were verified to have the three-dimensional profile characteristics of a protein spot. Those spots with volumes close to background level and dust particles with very large slopes and small areas were eliminated. The total number of protein spots detected on the master gel was 2556. Three hundred and ninety seven (15%) spots were matched on at least 37 gels, and 1215 (46%) spots were matched on at least 25 gels. The following statistical analyses were restricted to the 397 high-quality spots matched on 75% of the gels. These high-quality spots were chosen to focus on spots that did not require warping or imputing of any missing data. Future data analysis may help determine if warping and data imputation can expand high-quality gel proteomic data. The latest version of DeCyder contains the ability to warp gels to potentially add missing data. These additions may provide additional spots that can be studied as high-quality, but this type of data manipulation may also create skewed expression values, as the results depend on the type of postrun imputation model that is utilized [35].

Our previous study with technical replicates of one human plasma sample [13] had 42% of the spots matched on 8 of 12 gels. The addition of biological samples from different subjects at varying time points added to the complexity of the current dataset and reduced the matching accuracy. While most of the decrease in the matching accuracy is expected to
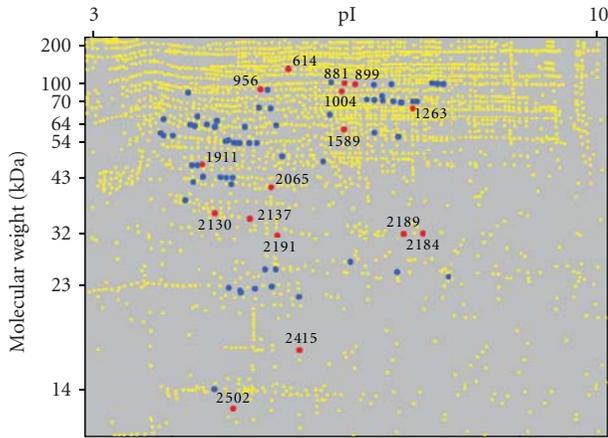
FIGURE 1: The spatial distribution of proteins spots (yellow dots) detected in human plasma by 2-D DIGE. Identified proteins (blue dots) and those showing differences between the theoretical and observed molecular weights (numbered red dots) are highlighted.

stem from the biological complexity of the experiment, part of it may be attributed to the larger number of gels, which inevitably increases the expected experimental variation. A study of five commercial software programs showed that an average of only 3% of the total analysis time was automated as opposed to manual, and as the number of gels increased, the percentage of automatically generated correct matches was dramatically reduced [36]. Taken together, these studies suggest that improved spot detection and matching software and algorithms are needed to increase the quality of spot matching. One such study was accomplished that created algorithms to improve spot matching with an integrated approach using hierarchical-based and optimization-based methods [37].

3.3. Differential Expression and Protein Identifications.

The pairwise comparisons among the 33 samples identified over 1400 spots with P-value <.05 and fold-change > 1.5. Down selection using manual inspection eliminated most of these spots (due to three-dimensional profile characteristics not representative of protein or because of insufficient representation of the spot on enough gels) resulting in 427 spots of further interest. The majority of the spots that did not pass manual verification were similar to background levels, lacking visual characteristics of protein spots. The sensitive detection parameters used in this study, while allowing for the detection of low abundance proteins, results in increased detection of artifacts that require manual verification.

Of the 427 spots with differential abundance, those exhibiting the greatest differences in abundance levels were further characterized, and 78 proteins spots were identified by mass spectrometry. The identified proteins are listed in Table 2, along with the theoretical pI and molecular weight calculated from the full-length amino acid sequence of each protein. Figure 1 depicts the spatial distribution of the protein spots detected in human plasma by 2-D DIGE. Identified proteins are denoted by blue dots.

Sixteen proteins (red dots) were found to have differences between the theoretical and observed molecular weights. The discrepancies are potentially due to posttranslational modifications or experimental processing; however, since all samples were treated identically, posttranslational modification is more likely. For example, spots 2189 and 2184, both identified as complement component C4A, were found to be statistically significant with at least a 1.5-fold difference between individuals. Complement component C4A has a theoretical molecular weight of 192.8 kDa; yet the protein spots identified indicate an approximate 32 kDa fragment. Since only C-terminal peptides were detected by mass spectrometry, the protein spot likely corresponds to the active Complement C4c fragment (mw = 33 kDa), which is a known cleavage product of Complement C4A [38, 39]. Variability in the amount of Complement C4c fragment between individuals could be a reflection of immune status, which may be a considerable variable when comparing human clinical subjects.

3.4. Spotwise Variation.

The distribution of the SLA was consistent across the gels. The spot-wise SD values of the SLA for the 397 high-quality spots, when considering all samples, ranged from 0.04 to 0.53, with a median of 0.10. When broken down separately by subject, the range was 0.0002 to 0.50, with 0.06 as the median, reflecting the lower variation of time-within-subjects than variation between the different subjects. Both sets of values represented an increase from the spot-wise SDs observed among technical replicates of the same human plasma sample [13], where the maximum was 0.20 and the median 0.04. In the previous TVS work [13], the CV values of the SA had a median of 10% and a maximum of 42%. Here, the range of the spot-wise CVs was 10% to 93%, with a median CV of 23%. The higher CVs of the present study reflect the additional complexity due to the heterogeneity of the samples from multiple human subjects. These results are comparable to the recently reported 6% (min), 108% (max), and 19% (median) CVs found in a 2-D DIGE study of normal liver samples from ten human subjects [40]. Here, about 90% of the spots had less than 40% CV, and only 21 spots (5% of the 397 high-quality spots) had higher than 50% CV. The spots are likely not good biomarker candidates due to their high individual or longitudinal variability. These spots showing relatively high variation may correspond to a single isoform of individual proteins and do not represent all isoforms of any given protein. Notably, several of the proteins identified (Albumin, Transferrin, Haptoglobin, IgG, and IgA) are proteins removed by the Top-6 depletion process [12], which was subsequently found to result in variability when processing multiple samples in series. In future studies, column equilibration steps are recommended between samples to reduce this variability and ensure more complete depletion of high-abundant proteins.

In summary, the majority of the spots had small enough CV to indicate that the corresponding protein expressions were relatively constant across individuals, and thus could be potentially used as biomarkers. The minimum, maximum, and median CVs, when calculated separately for the subjects

TABLE 1: Frequency distribution of the variance component estimates.

% contribution to total variance	Subject		Time in subject		Error	
	(a)	(b)	(a)	(b)	(a)	(b)
0–10	6.31	6.31	63.89	63.89	6.57	6.57
10–20	9.34	15.66	17.68	81.57	17.17	23.74
20–30	12.89	28.50	7.83	89.39	16.67	40.40
30–40	14.14	42.68	4.29	93.69	9.34	49.75
40–50	13.63	56.31	2.27	95.96	10.86	60.61
50–60	11.36	67.68	3.03	98.99	11.87	72.47
60–70	11.36	79.04	0.51	99.49	10.37	82.83
70–80	11.61	90.66	0.50	100.00	8.08	90.91
80–90	7.83	98.48			4.80	95.71
90–100	1.51	100.00			4.29	100.00

The components of subject (σ_s), time-within-subject (σ_t), and random error (σ_e) are shown separately as (a) the percentage of spots and (b) the cumulative percentage of spots with contribution to the total variance indicated in the first column.

were 0.05%, 131%, and 14%, respectively. Over 95% of the CVs were below 35%, indicating that for most subjects, the variation over the three timepoints was comparable to the experimental variation in the previously published TVS data [13].

3.5. Statistical Normalization, Gel and Dye Effect Removal, and Variance Decomposition. The SLA values were further normalized as explained in the methods. The effect of the normalization on the results is addressed as appropriate in the following sections. The F-tests for the analysis of variance calculations corresponding to (1) indicated significant gel (P-value $< 2.2e-16$), dye (P-value $< 3.2e-16$), and gel-dye interaction (P-value $< 2.2e-16$) effects. The residual diagnostic plots did not reveal major departures from the assumptions, thus indicating the validity of the model. Similar analyses using the normalized SLA resulted in slightly higher P-values (gel effect P-value $< 2.2e-16$, dye effect P-value.003, gel-dye interaction P-value $< 3.0e-09$) but were consistent with the conclusions based on the calculations using the SLA.

The standard deviations corresponding to the random variance component estimates from the mixed-effects model (Figure 2) show the relative contribution of the three components at each of the spots matched on 75% of the gels. Overall, the time-within-subject component was found to have the smallest contribution to the total variance, while subject-related variation had the highest. The corresponding frequency distributions of the three variance components (Table 1) confirm that for most spots (89%) the contribution of the time component was less than 30% of the total variance. Only 5% of the spots had 70% to 80% of their variation explained by the time component, and no spot had the time component greater than 80% of the total variance. For 21% of the spots, the contribution of the subject component comprised over 70% of the total variance. For about 44% of the spots, the contribution of the subject component represented over 50% of the total variation.

To further elucidate the contributing factors involved in spot variance, we performed a meta-analysis on these data.
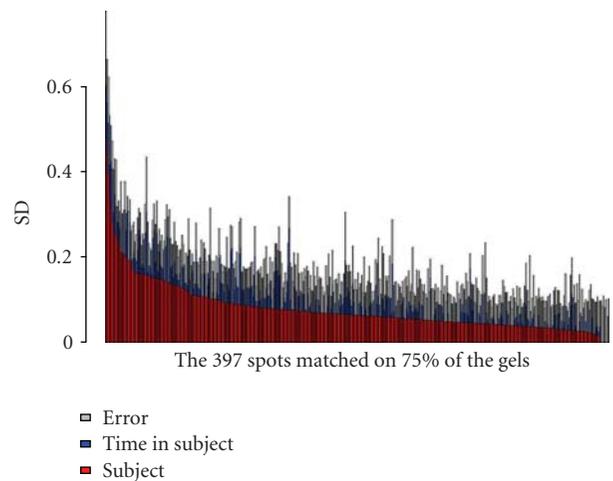
FIGURE 2: The subject (σ_s), time-within-subject (σ_t), and random error (σ_e) variance component estimates (on SD scale) for the 397 protein spots matched on at least 75% of the gels, ordered by the magnitude of the subject component.

Essentially, all the variances for the 397 high-quality spots were summed and the total variance that could be explained by the sum of the spot-wise subject, time within-subject, and error components was determined, respectively. A pooled estimate of the variance components was obtained by taking the average of the corresponding variance components over the spots. When aggregating the total variance over all the spots matched on at least 75% of the gels, the sum of the subject components explained 59% of the total variation and the sum of the time-within-subject components explained 12% of the total variation. The average subject variance component across the spots was 0.0097 (corresponding to σ_s = 0.098 on the Sd scale), and the average time-within-subject variance component was 0.0019 (σ_t = 0.044).

3.6. Multivariate Analysis of Expression Patterns. The EDA module of DeCyder 2D was used to visually display the

Table 2: Variable proteins identified from human plasma.

Protein Number	Protein Identity	Accession Number	mw[a]	pI[a]	Gene
614a	clusterin	IPI00400826	57.8	6.25	CLU
614b	alpha-2-macroglobulin precursor	IPI00478003	163.3	6	A2M
835a	alpha-2-macroglobulin precursor	IPI00478003	163.3	6	A2M
835b	Complement C3 precursor	IPI00783987	187.1	6.02	C3
849	Plasminogen	IPI00019580	90.6	7.04	PLG
856	Plasminogen	IPI00019580	90.6	7.04	PLG
881a	complement factor B preproprotein	IPI00019591	85.5	6.67	CFB
881b	complement protein C7 precursor	IPI00296608	93.5	6.09	C7
881c	Complement C3 precursor	IPI00783987	187.1	6.02	C3
884	Plasminogen	IPI00019580	90.6	7.04	PLG
893	complement factor B preproprotein	IPI00019591	85.5	6.67	CFB
899a	complement factor B preproprotein	IPI00019591	85.5	6.67	CFB
899b	complement protein C7 precursor	IPI00296608	93.5	6.09	C7
899c	Complement C3 precursor	IPI00783987	187.1	6.02	C3
910a	fibrinogen gamma	IPI00219713	49.5	5.7	FGG
910b	complement factor B preproprotein	IPI00019591	85.5	6.67	CFB
956a	complement component 1, r subcomponent	IPI00296165	80.2	5.89	C1R
956b	complement component C4A	IPI00032258	192.8	6.66	C4A
963	complement component 1, r subcomponent	IPI00296165	80.2	5.89	C1R
1002	complement component 1,s subcomponent	IPI00017696	76.7	4.86	C1S
1004a	gelsolin	IPI00646773	80.6	5.58	GSN
1004b	complement component 2	IPI00303963	83.3	7.23	C2
1004c	complement factor B preproprotein	IPI00019591	85.5	6.67	CFB
1004d	complement protein C7 precursor	IPI00296608	93.5	6.09	C7
1004e	alpha-2-macroglobulin precursor	IPI00478003	163.3	6	A2M
1004f	Complement C3 precursor	IPI00783987	187.1	6.02	C3
1004g	complement component C4A	IPI00032258	192.8	6.66	C4A
1027	transferrin	IPI00022463	77	6.81	TF
1110	IGHM protein	IPI00828205	65.3	8.1	IGHM
1113a	IGHM protein	IPI00828205	65.3	8.1	IGHM
1113b	transferrin	IPI00022463	77	6.81	TF
1128a	IGHM protein	IPI00828205	65.3	8.1	IGHM
1128b	transferrin	IPI00022463	77	6.81	TF
1129a	histidine-rich glycoprotein precursor	IPI00022371	59.6	7.09	HRG
1129b	coagulation factor XII precursor	IPI00019581	67.5	7.94	F12
1129c	transferrin	IPI00022463	77	6.81	TF
1142a	histidine-rich glycoprotein precursor	IPI00022371	59.6	7.09	HRG
1142b	transferrin	IPI00022463	77	6.81	TF
1156	transferrin	IPI00022463	77	6.81	TF
1185	transferrin	IPI00022463	77	6.81	TF
1254	hemopexin precursor	IPI00022488	51.5	6.57	HPX
1263a	histidine-rich glycoprotein precursor	IPI00022371	59.6	7.09	HRG
1263b	Complement C3 precursor	IPI00783987	187.1	6.02	C3
1276a	hemopexin precursor	IPI00022488	51.5	6.57	HPX
1276b	Heparin cofactor II precursor	IPI00292950	57.1	6.41	HCF2
1276c	peptidoglycan recognition protein L precursor	IPI00163207	62.2	7.25	PGLYRP
1382	kininogen	IPI00215894	47.9	6.29	KNG
1394	albumin	IPI00745872	69.1	5.85	ALB
1456	alpha-1-antichymotrypsin precursor	IPI00550991	45.5	5.32	AACT
1471a	immunoglobulin alpha-1 heavy chain	IPI00166866	37.6	6.06	IGHA1
1471b	kininogen	IPI00215894	47.9	6.29	KNG

TABLE 2: Continued.

Protein Number	Protein Identity	Accession Number	mw^a	pI^a	Gene
1471c	antithrombin III	IPI00032179	52.6	6.32	AT3
1525	Vitronectin precursor	IPI00298971	54.3	5.55	VTN
1526a	kininogen	IPI00215894	47.9	6.29	KNG
1526b	Angiotensinogen	IPI00032220	53.2	5.78	AGT
1555	Vitronectin precursor	IPI00298971	54.3	5.55	VTN
1558	immunoglobulin alpha-2 heavy chain	IPI00641229	36.4	5.71	IGH
1568a	kininogen	IPI00215894	47.9	6.29	KNG
1568b	antithrombin III	IPI00032179	52.6	6.32	AT3
1568c	Angiotensinogen	IPI00032220	53.2	5.78	AGT
1577	Alpha-2-HS-glycoprotein	IPI00022431	39.3	5.43	AHSG
1589a	immunoglobulin alpha-1 heavy chain	IPI00166866	37.6	6.06	IGHA1
1589b	apolipoprotein H precursor	IPI00298828	38.3	8.34	APOH
1589c	prepro-plasma carboxypeptidase B	IPI00329775	48.4	7.61	pCPB
1589d	fibrinogen beta chain	IPI00298497	55.9	8.54	FGB
1589e	alpha-2-macroglobulin precursor	IPI00478003	163.3	6	A2M
1616a	apolipoprotein H precursor	IPI00298828	38.3	8.34	APOH
1616b	fibrinogen beta chain	IPI00298497	55.9	8.54	FGB
1626	Alpha-2-HS-glycoprotein	IPI00022431	39.3	5.43	AHSG
1648	fibrinogen beta chain	IPI00298497	55.9	8.54	FGB
1650a	apolipoprotein D	IPI00006662	28	5.14	APOD
1650b	Alpha-2-HS-glycoprotein	IPI00022431	39.3	5.43	AHSG
1650c	alpha-1-antichymotrypsin precursor	IPI00550991	45.5	5.32	AACT
1650d	fibrinogen gamma	IPI00219713	49.5	5.7	FGG
1652	Alpha-2-HS-glycoprotein	IPI00022431	39.3	5.43	AHSG
1725	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1731	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1740	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1741	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1744	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1749	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1752	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1843a	pigment epithelial-differentiating factor	IPI00006114	46.3	5.84	PEDF
1843b	fibrinogen gamma	IPI00219713	49.5	5.7	FGG
1898	vitamin D-binding protein precursor	IPI00742696	52.9	5.32	GC
1911a	serum paraoxonase	IPI00218732	37.8	4.96	PON
1911b	fibrinogen gamma	IPI00219713	49.5	5.7	FGG
1911c	Complement C3 precursor	IPI00783987	187.1	6.02	C3
1918	serum paraoxonase	IPI00218732	37.8	4.96	PON
1925a	serum paraoxonase	IPI00218732	37.8	4.96	PON
1925b	fibrinogen gamma	IPI00219713	49.5	5.7	FGG
1985a	serum paraoxonase	IPI00218732	37.8	4.96	PON
1985b	haptoglobin	IPI00641737	45.2	6.13	HP
1986a	apolipoprotein A-IV precursor	IPI00304273	43.4	5.22	APOA4
1986b	haptoglobin	IPI00641737	45.2	6.13	HP
1986c	serum paraoxonase	IPI00218732	37.8	4.96	PON
1998	apolipoprotein A-IV precursor	IPI00304273	43.4	5.22	APOA4
2008	apolipoprotein A-IV precursor	IPI00304273	43.4	5.22	APOA4
2029	alpha-2-glycoprotein 1	IPI00166729	34.3	5.71	AZGP1
2030a	apolipoprotein A-IV precursor	IPI00304273	43.4	5.22	APOA4
2030b	haptoglobin	IPI00641737	45.2	6.13	HP
2065a	haptoglobin	IPI00641737	45.2	6.13	HP

TABLE 2: Continued.

Protein Number	Protein Identity	Accession Number	mw[a]	pI[a]	Gene
2065b	Complement factor I	IPI00291867	65.8	7.72	CFI
2095a	alpha-1-antichymotrypsin precursor	IPI00550991	45.5	5.32	AACT
2095b	clusterin	IPI00400826	57.8	6.25	CLU
2130a	Proapolipoprotein	IPI00021841	29	5.45	APOA1
2130b	Complement factor I	IPI00291867	65.8	7.72	CFI
2137	clusterin	IPI00400826	57.8	6.25	CLU
2184	complement component C4A	IPI00032258	192.8	6.66	C4A
2189	complement component C4A	IPI00032258	192.8	6.66	C4A
2191	transthyretin	IPI00022432	15.9	5.5	TTR
2236	immunoglobulin kappa light chain	IPI00784070	26	8.16	IGKC
2259	amyloid P component	IPI00022391	25.4	6.1	APCS
2260	amyloid P component	IPI00022391	25.4	6.1	APCS
2272a	lambda-chain precursor	IPI00154742	24.7	7.54	IGL
2272b	immunoglobulin kappa light chain	IPI00784070	26	8.16	IGKC
2284	immunoglobulin kappa light chain	IPI00784070	26	8.16	IGKC
2314	Proapolipoprotein	IPI00021841	29	5.45	APOA1
2325	Proapolipoprotein	IPI00021841	29	5.45	APOA1
2326	Proapolipoprotein	IPI00021841	29	5.45	APOA1
2338	Proapolipoprotein	IPI00021841	29	5.45	APOA1
2346	plasma glutathione peroxidase	IPI00026199	16.7	8.93	GPx-P
2415	haptoglobin	IPI00641737	45.2	6.13	HP
2468	transthyretin	IPI00022432	15.9	5.5	TTR
2520	haptoglobin	IPI00641737	45.2	6.13	HP

[a]Theoretical molecular weight (mw) in kDa and isoelectric point (pI) values.

results of the current study and the previous TVS study [13] (Figure 3). The multivariate expression profiles of the samples across 328 spots that were matched on >75% of the spot maps from both studies were transformed into the principal component basis and the projection of the samples onto the first two principal components displayed (Figure 3). The tight scatter of the samples from the TVS (encircled in black) indicates the small magnitude of the experimental variability when analyzing technical replicates of the same human sample. The magnitude of the longitudinal variation exceeded the technical variation, as evidenced by the larger scatter of the sample points of a given subject at the different time points. For example, the two red and green ellipses (Figure 3) highlight the longitudinal variation for subjects 1 and 11, respectively. The differential scattering of the samples, from the subjects into varying regions of the principal components plot, indicates that the subject-to-subject variation exceeded the longitudinal variation within subjects.

Hierarchical clustering was used to group the 33 samples based on the similarity of their protein expression profiles along the 397 high-quality spots that were matched on >75% of the gels (Figure 4). Clustering was performed in on the proteins and experimental samples, using Euclidean distance and average linkage to define similarity. For all subjects, the first clustering step placed the three samples of the given subject into one cluster. Samples of the same subject collected at the three time points were most similar to each
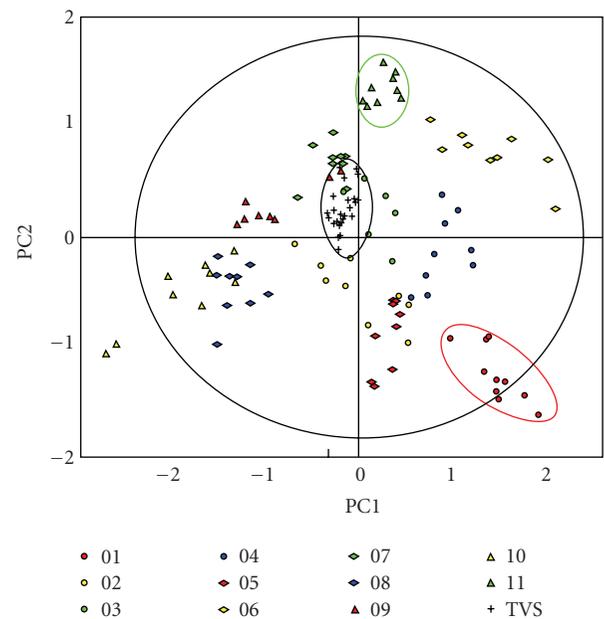
FIGURE 3: Principal component analysis of the 33 samples from the present study and the 8 replicates from the previous Technical Variation Study (TVS) [9], color-coded according to the legend, projected onto the first two principal components. Ellipses highlighting subjects 1 (red), 11 (green), and the TVS (black) are added for illustrative purposes only.
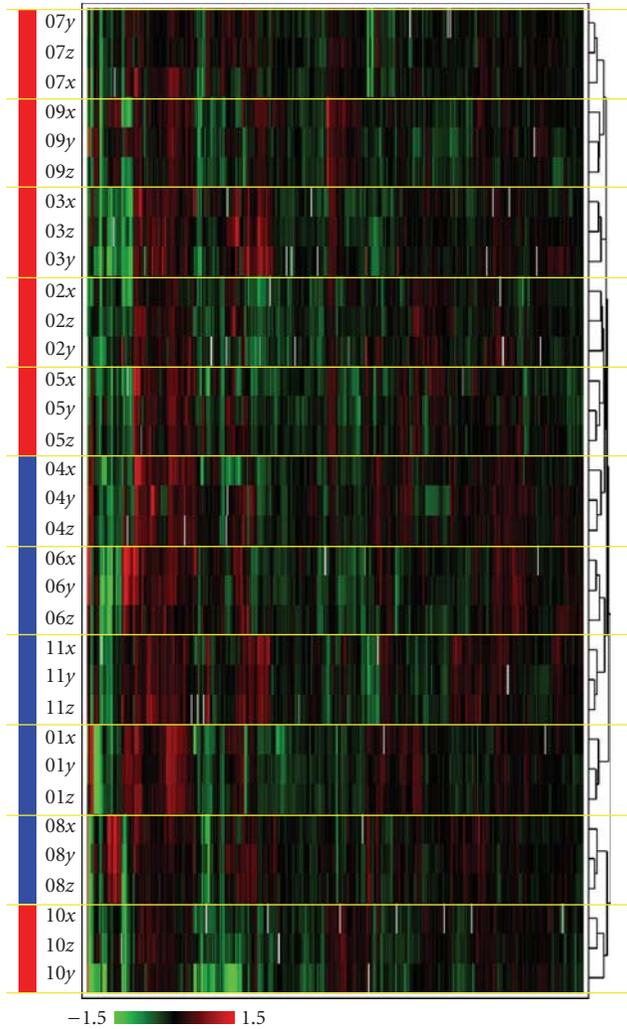
FIGURE 4: Hierarchical clustering of the 33 samples (*y*-axis) based on the abundance of the 397 high-quality protein spots on the *x*-axis, using Euclidean distance and average linkage. The samples are in SubjectNumberTime format, where SubjectNumber ranges from 01 to 11, and the Time values {*x*, *y*, and *z* } correspond to {T_1, T_2, and T_3}. The intensities range from −1.5-fold change (bright green) to 1.5-fold change (bright red). The dendrogram on the right indicates the order of the sample grouping, with more similar samples being grouped together first. The color band on the left shows the genders of the samples, with red for females, and blue for males.

other, as evidenced by the succession of self-similar bands of three rows (highlighted by the yellow lines in Figure 4). The clustering also shows a general trend of clustering the samples based on gender appeared (highlighted by the blue and red bars for males and females, resp., in Figure 4).

3.7. Gender Effects. In addition to the results seen in the hierarchical clustering (Figure 4), after fitting the mixed-effects model to the residuals from the SLA at the spots
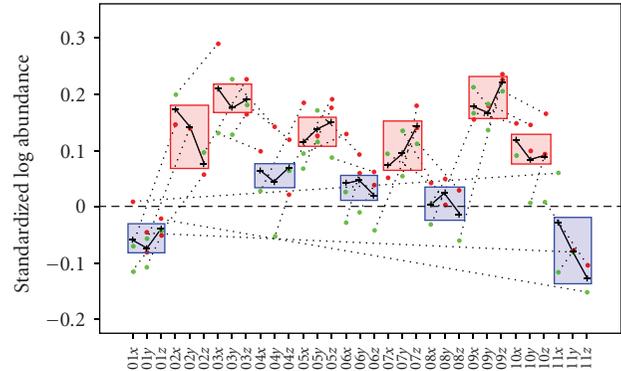
FIGURE 5: Expression data for alpha-2-HS-glycoprotein, with an average increase of 1.49-fold between the female and male groups, and FDR-adjusted gender-effect *P*-value =.055. The samples are in SubjectNumberTime format, where SubjectNumber ranges from 01 to 11, and the Time values {*x*, *y*, and *z* } correspond to {T_1, T_2, and T_3}. The annotations indicate the gels (numbers) and the dyes (red for Cy5, green for Cy3) corresponding to the samples. Dotted lines connect samples multiplexed on the same gel. Crosses indicate sample averages over the technical replicates. The solid line connects all sample averages. Boxes around the three Time values for each SubjectNumber highlight male and female genders (blue and red respectively) added for illustrative purposes only.

matched on 75% of the gels, 17 spots showed gender-effect *P*-value <.01. None of these spots were found to be significant (*P*-value <.01) after the False Discovery Rate (FDR) method [41] for multiple comparisons was applied suggesting that larger numbers of samples are needed to validate gender differences in the human plasma proteome. Despite the lack of statistically significant data on gender differences, trends in this dataset suggest that future, larger datasets might enable the differentiation of protein expression levels due to gender. One spot, 1659, had FDR-adjusted gender-effect *P*-value equal to.055 with a 1.49-fold-change between the male and female groups (Figure 5). Five additional spots (466, 1626 alpha-2-HS-glycoprotein, 1650 alpha-2-HS-glycoprotein, 1652 alpha-2-HS-glycoprotein, 1678) had adjusted *P*-values of.11. The results were similar when fitting the same model to the residuals from the statistically normalized SLA, albeit with *P*-values that slightly exceeded their corresponding values based on the SLA. Three of the spots exhibiting gender effects were identified as alpha-2-HS-glycoprotein, which has been shown to vary between males and females. The concentration of alpha-2-HS-glycoprotein has been found to undergo a progressive age-related decrease in women, while men show no noticeable change [36].

The removal of the dye and gel effects using the model in (1) proved to be a beneficial preprocessing step. Without this step, when the mixed-effect model was fit to the original SLA, the smallest FDR-adjusted gender-effect *P*-value was.19 (spot 1659). When the same model was fit to the statistically normalized SLA, the smallest FDR-adjusted *P*-value was also.19 (spot 1659). The statistical normalization improved the quality of the data slightly, but it did not reduce dramatically the observed *P*-values. On the other
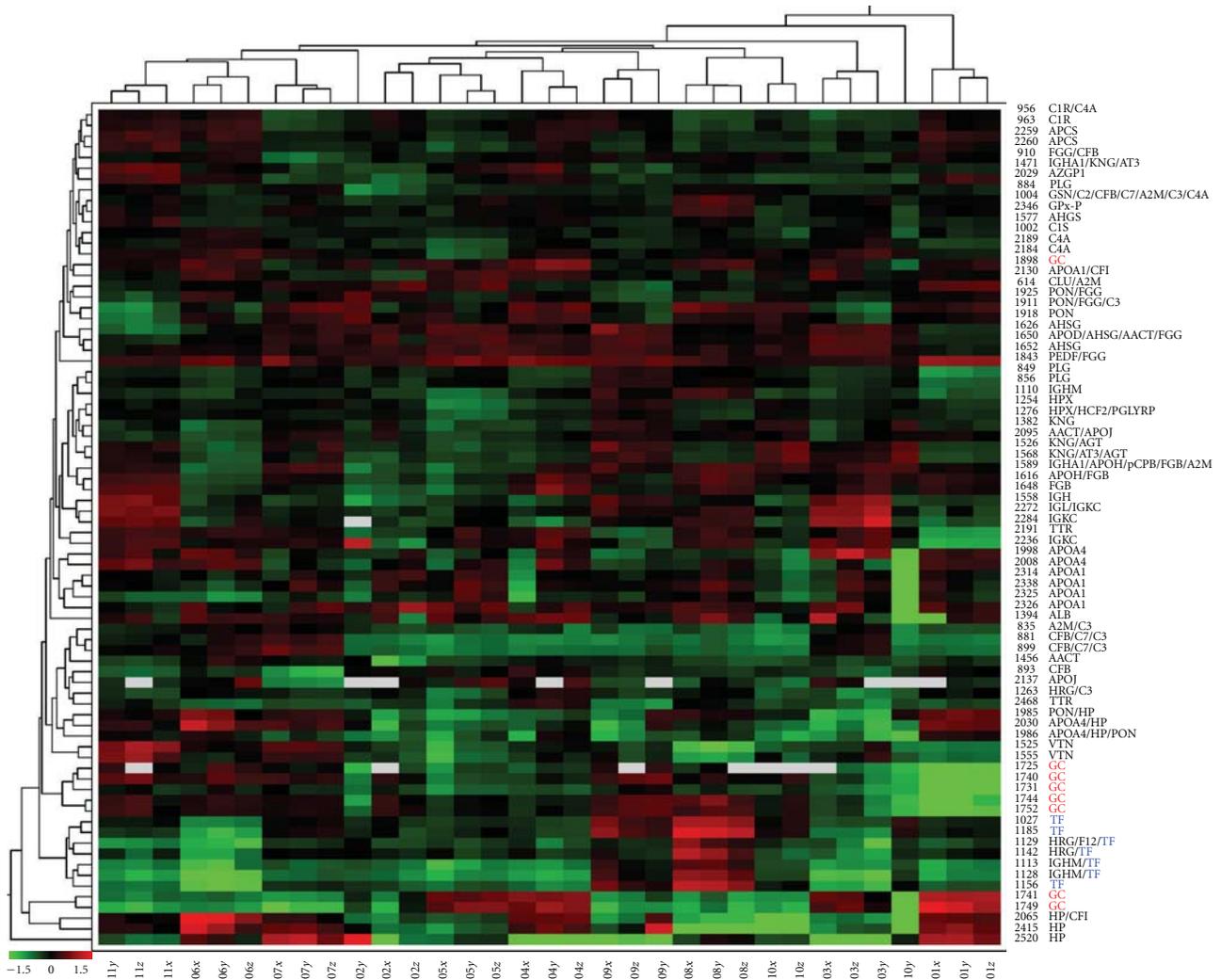
FIGURE 6: Hierarchical clustering of the 33 samples (*x*-axis) based on the abundance of the 78-identified protein spots on the *y*-axis, using Euclidean distance metrics and average linkage methods. The samples are in SubjectNumberTime format, where SubjectNumber ranges from 01 to 11, and the Time values (*x*, *y*, and *z*) correspond to (T_1, T_2, and T_3). The intensities range from -1.5 (bright green) to 1.5 (bright red). The dendrogram on the top indicates the order of the sample grouping, with samples corresponding to the lower leaves being grouped together first. Similarly, the dendrogram on the left indicates the ordering of the protein spots. All transferrin (TF) and vitamin D-binding protein (GC) identifications are highlighted in blue and red, respectively.

hand, pooling the information across the gels to remove the common dye and gel effects strengthened the signal and reduced markedly the FDR-adjusted *P*-values. As explained in the previous paragraph, for spot 1659, the new *P*-value was close to .05.

3.8. Multivariate Analysis of Identified Proteins.

Hierarchical clustering of the identified proteins (Figure 6) was conducted using the Euclidean distance metric and average linkage methods. For all subjects, other than subject 10, the first clustering step placed the three samples of that subject into one cluster. Subject 10 had two time points grouped together (*x* and *z*) with the third point (*y*) separated by Subject 3. Because the 78 identified proteins were the most differential

between time and subjects, it is not unexpected to see clustering results that may not perfectly align all subjects or time points. For example, all protein spots identified as transferrin (TF) clustered together due to their similar expression patterns, while the vitamin D-binding protein (GC) spots were found in multiple clusters due to differences in expression patterns between the individuals. Multiple proteins may cluster together due to coregulation and similar functions, and in the case of APOA4 and APOA1 (Figure 6), coregulation has been reported [37, 42].

4. Conclusion

Statistical analysis of a 2-D DIGE experiment involving triplicate plasma samples from eleven human subjects taken

at three time points separated by several weeks demonstrated that the subject-to-subject variation exceeded the time-within-subject variation. The variation in the human plasma proteome reported here was greater than a previous technical variation study wherein one plasma sample was processed multiple times [13].

Here, for 70% of the high-quality protein spots, the coefficient of variation of the SLA was less than 30% across all subjects and time points, thus indicating that the baseline expression levels of those proteins are relatively stable in the population represented by the subjects in this study. Only 21 spots had larger than 50% CV, suggesting that these protein isoforms should be avoided as biomarker candidates. Many of these protein spots represent medium to high abundance plasma proteins. Since they are higher abundance, they might bias LC/MS datasets, but since the total number of these spots relative to the total plasma proteome is small, their total influence on a sample is likely also small. In addition, protein spots with gender-related differences should be considered separately for males and females. However, more thorough studies, including the use of a larger population set with additional time points over longer periods of time, are recommended to more fully address individual, longitudinal, and gender variability as related to biomarker discovery.

We noted that preprocessing the data by first removing the fixed effects of the gels and dyes was important in data analysis and improved the quality of the data. This step resulted in six protein spots showing a statistically significant gender effect at an FDR-adjusted 11% significance level. Without the preprocessing step, the smallest gender effect P-value was.19. While removing the gel and dye effects lead to stronger conclusions, the additional statistical normalization of the SLA had only marginal effects and did not alter the conclusions.

Spot matching confounds gel- and software-related protein differences with real biological effects. In the present study, we only considered spots that were matched on at least 75% of the gels. Spots with lower matching quality can be investigated separately, as they may correspond to proteins that are absent or have very low expression in certain individuals, but which may have biological significance. We envision that such studies will become more relevant as the field of personalized medicine matures, and as detection and matching algorithms continue to improve.

This study represents a first step toward quantitating the longitudinal and individual variation in the human plasma proteome, as measured on the 2-D DIGE platform. Interestingly, gender-related variations were also detected suggesting that gender variability should also be considered in biomarker discovery. Future, larger-scale experiments that include more subjects representative of various population segments, encompassing differences in ethnicity, age, gender, disease status, and other relevant factors, have the potential to define baseline proteomic similarities and differences in the human population, which will in turn facilitate improved biomarker discovery.

Acknowledgments

References

[1] D. Nedelkov, "Population proteomics: addressing protein diversity in humans," *Expert Review of Proteomics*, vol. 2, no. 3, pp. 315–324, 2005.

[2] H. Hermjakob, "The HUPO proteomics standards initiative—overcoming the fragmentation of proteomics data," *Proteomics*, vol. 1, no. S2, pp. 34–38, 2006.

[3] C. F. Taylor, "Minimum reporting requirements for proteomics: a MIAPE primer," *Proteomics*, vol. 1, no. S2, pp. 39–44, 2006.

[4] G. S. Omenn, "Exploring the human plasma proteome: editorial," *Proteomics*, vol. 5, no. 13, pp. 3223–3225, 2005.

[5] G. S. Omenn, Y.-K. Paik, and D. Speicher, "The HUPO plasma proteome project: a report from the Munich congress," *Proteomics*, vol. 6, no. 1, pp. 9–11, 2006.

[6] G. S. Omenn, D. J. States, M. Adamski, et al., "Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database," *Proteomics*, vol. 5, no. 13, pp. 3226–3245, 2005.

[7] K. Cottingham, "HUPO plasma proteome project: challenges and future directions," *Journal of Proteome Research*, vol. 5, no. 6, p. 1298, 2006.

[8] J. M. Jacobs, J. N. Adkins, W.-J. Qian, et al., "Utilizing human blood plasma for proteomic biomarker discovery," *Journal of Proteome Research*, vol. 4, no. 4, pp. 1073–1085, 2005.

[9] S. M. Hanash, S. J. Pitteri, and V. M. Faca, "Mining the plasma proteome for cancer biomarkers," *Nature*, vol. 452, no. 7187, pp. 571–579, 2008.

[10] M. J. Han and D. W. Speicher, "Microscale isoelectric focusing in solution: a method for comprehensive and quantitative proteome analysis using 1-D and 2-D DIGE combined with MicroSol IEF prefractionation," *Methods in Molecular Biology*, vol. 424, pp. 241–256, 2008.

[11] S. Y. Cho, E.-Y. Lee, H.-Y. Kim, et al., "Protein profiling of human plasma samples by two-dimensional electrophoresis," *Methods in Molecular Biology*, vol. 428, pp. 57–75, 2007.

[12] B. A. Chromy, A. D. Gonzales, J. Perkins, et al., "Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins," *Journal of Proteome Research*, vol. 3, no. 6, pp. 1120–1127, 2004.

[13] T. H. Corzett, I. K. Fodor, M. W. Choi, et al., "Statistical analysis of the experimental variation in the proteomic characterization of human plasma by two-dimensional difference gel electrophoresis," *Journal of Proteome Research*, vol. 5, no. 10, pp. 2611–2619, 2006.

[14] A. Alban, S. O. David, L. Bjorkesten, et al., "A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating

a pooled internal standard," *Proteomics*, vol. 3, no. 1, pp. 36–44, 2003.

[15] K. S. Lilley and D. B. Friedman, "All about DIGE: quantification technology for differential-display 2D-gel proteomics," *Expert Review of Proteomics*, vol. 1, no. 4, pp. 401–409, 2004.

[16] R. Marouga, S. David, and E. Hawkins, "The development of the DIGE system: 2D fluorescence difference gel analysis technology," *Analytical and Bioanalytical Chemistry*, vol. 382, no. 3, pp. 669–678, 2005.

[17] L. A. Echan, H.-Y. Tang, N. Ali-Khan, K. Lee, and D. W. Speicher, "Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma," *Proteomics*, vol. 5, no. 13, pp. 3292–3303, 2005.

[18] R. C. Mahnke, T. H. Corzett, S. L. McCutchen-Maloney, and B. A. Chromy, "An integrated proteomic workflow for two-dimensional differential gel electrophoresis and robotic spot picking," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2093–2097, 2006.

[19] W. N. Venables and D. M. Smith, *An Introduction to R: Notes on R, A Programming Environment for Data Analysis and Graphics, v.2.0.1*, Network Theory, Bristol, UK, 2004.

[20] B. A. Chromy, J. Perkins, J. L. Heidbrink, et al., "Proteomic characterization of host response to Yersinia pestis and near neighbors," *Biochemical and Biophysical Research Communications*, vol. 320, no. 2, pp. 474–479, 2004.

[21] N. Bolshakova and F. Azuaje, "Cluster validation techniques for genome expression data," *Signal Processing*, vol. 83, no. 4, pp. 825–833, 2003.

[22] J. C. Dunn, "Well separated clusters and optimal fuzzy partitions," *The Journal on Systemics, Cybernetics and Informatics*, vol. 4, pp. 95–104, 1974.

[23] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14863–14868, 1998.

[24] L. Kaufmann and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.

[25] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, vol. 38, no. 6, pp. 1409–1438, 1958.

[26] R. Tibshirani, G. Walther, and T. Hastie, *Estimating the Number of Clusters in a Dataset via the Gap Statistic*, Department of Biostatistics, Stanford University, Stanford, Calif, USA, 2000.

[27] L. Eriksson, N. Kettaneh-Wold, J. Trygg, and S. Wold, *Multi- and Megavariate Data Analysis*, vol. 533, Umetrics Academy, Umeå, Sweden, 2001.

[28] I. T. Jolliffe, *Principal Component Analysis*, Springer, Berlin, Germany, 2002.

[29] "Umetrics SIMCA,10.5," Sweden.

[30] H. Wold, "Estimation of principal components and related models by iterative least squares," in *Multivariate Analysis*, P. Krishnaiah, Ed., pp. 391–420, Academic Press, New York, NY, USA, 1966.

[31] J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer, New York, NY, USA, 2000.

[32] R. O. Kuehl, *Design of Experiments: Statistical Principles of Research Design and Analysis*, Duxbury Press, Belmont, Calif, USA, 2nd edition, 2000.

[33] X. Cui and G. A. Churchill, "Statistical tests for differential expression in cDNA microarray experiments," *Genome Biology*, vol. 4, no. 4, article 210, 2003.

[34] I. K. Fodor, D. O. Nelson, M. Alegria-Hartman, et al., "Statistical challenges in the analysis of two-dimensional difference gel electrophoresis experiments using DeCyder™," *Bioinformatics*, vol. 21, no. 19, pp. 3733–3740, 2005.

[35] R. Pedreschi, M. L. A. T. M. Hertog, S. C. Carpentier, et al., "Treatment of missing values for multivariate statistical analysis of gel-based proteomics data," *Proteomics*, vol. 8, no. 7, pp. 1371–1383, 2008.

[36] I. R. Dickson, M. Bagga, and C. R. Paterson, "Variations in the serum concentration and urine excretion of α2HS-glycoprotein, a bone-related protein, in normal individuals and in patients with osteogenesis imperfecta," *Calcified Tissue International*, vol. 35, no. 1, pp. 16–20, 1983.

[37] O. Schamaun, B. Olaisen, B. Mevag, et al., "The two apolipoprotein loci apoA-I and apoA-IV are closely linked in man," *Human Genetics*, vol. 68, no. 2, pp. 181–184, 1984.

[38] I. Gigli, I. von Zabern, and R. R. Porter, "The isolation and structure of C4, the fourth component of human complement," *Biochemical Journal*, vol. 165, no. 3, pp. 439–446, 1977.

[39] E. M. Press and J. Gagnon, "Human complement component C4. Structural studies on the fragments derived from C4b by cleavage with C3b inactivator," *Biochemical Journal*, vol. 199, no. 2, pp. 351–357, 1981.

[40] X. Zhang, Y. Guo, Y. Song, et al., "Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis," *Proteomics*, vol. 6, no. 19, pp. 5260–5268, 2006.

[41] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society Series B*, vol. 57, no. 1, pp. 289–300, 1995.

[42] L. Vergnes, T. Taniguchi, K. Omori, M. M. Zakin, and A. Ochoa, "The apolipoprotein A-I/C-III/A-IV gene cluster: ApoC-III and ApoA-IV expression is regulated by two common enhancers," *Biochimica et Biophysica Acta*, vol. 1348, no. 3, pp. 299–310, 1997.

Research Article

Correctness of Protein Identifications of *Bacillus subtilis* Proteome with the Indication on Potential False Positive Peptides Supported by Predictions of Their Retention Times

Katarzyna Macur,[1] Tomasz Bączek,[1] Roman Kaliszan,[2] Caterina Temporini,[3] Federica Corana,[4] Gabriella Massolini,[3] Jolanta Grzenkowicz-Wydra,[5] and Michał Obuchowski[6]

[1] *Department of Pharmaceutical Chemistry, Medical University of Gdańsk, 80-416 Gdańsk, Poland*
[2] *Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland*
[3] *Department of Pharmaceutical Chemistry, University of Pavia, 27100 Pavia, Italy*
[4] *Centro Grandi Strumenti, Università degli Studi di Pavia, 27100 Pavia, Italy*
[5] *Pomeranian Science and Technology Park, Gdynia Innovation Centre, Gdynia, Poland*
[6] *Laboratory of Molecular Bacteriology, Department of Medical Biotechnology, Intercollegiate Faculty of Biotechnology, Medical University of Gdańsk, 80-211 Gdańsk, Poland*

Correspondence should be addressed to Tomasz Bączek, tbaczek@amg.gda.pl

The predictive capability of the retention time prediction model based on quantitative structure-retention relationships (QSRR) was tested. QSRR model was derived with the use of set of peptides identified with the highest scores and originated from 8 known proteins annotated as model ones. The predictive ability of the QSRR model was verified with the use of a *Bacillus subtilis* proteome digest after separation and identification of the peptides by LC-ESI-MS/MS. That ability was tested with three sets of testing peptides assigned to the proteins identified with different levels of confidence. First, the set of peptides identified with the highest scores achieved in the search were considered. Hence, proteins identified on the basis of more than one peptide were taken into account. Furthermore, proteins identified on the basis of just one peptide were also considered and, depending on the possessed scores, both above and below the assumed threshold, were analyzed in two separated sets. The QSRR approach was applied as the additional constraint in proteomic research verifying results of MS/MS ion search and confirming the correctness of the peptides identifications along with the indication of the potential false positives.

1. Introduction

Liquid chromatography (LC) combined with tandem mass spectrometry (MS/MS) plays an essential role in the field of protein research. In this technique, proteins and peptides are separated with the use of liquid chromatography methods and then identified by tandem mass spectrometry analysis. Thanks to high resolution, accuracy, and sensitivity of LC-MS/MS systems, equipped with sophisticated techniques of fragmentation, not only can simple proteins be directly investigated, but also research on the level of whole proteomes became possible [1]. However, proteins/peptides

identification from biological matrices is still an analytical challenge because of the great complexity of the samples, enormous concentration ranges of the occurring proteins and lack of proper standards. It all makes an exact and precise peptide or protein identification and, consequently, proteome coverage limited [2].

Proteomic research requires also higher throughput of the protein identification in LC-MS/MS. Peptide identification in MS/MS is based on matching to parent ion m/z and m/z values of daughter ions. This procedure allows to assign an identification confidence for this particular peptide, which contributes independently to the overall

confidence of the protein identification. One of the most commonly applied method for protein definition in complex samples relies on correlation algorithm Sequest proposed by Yates and coworkers [3–6]. This algorithm matches the investigated peptide tandem mass spectrometry data with proper data from protein database. To increase reliability of the identification, several statistic parameters have been considered. First, the difference between the normalized cross-correlation functions for the first and second ranked results (ΔC_n) is applied to indicate a correctly selected peptide sequence. The other criteria are cross-correlation score between the observed peptide fragment mass spectrum and the theoretically predicted one (X_{corr}), the preliminary score based on the number of ions in the MS/MS spectrum that match the experimental data (S_p), the rank of the certain match during the preliminary scoring (RS_p), and the ions value (I) describing how many of the observed ions match the theoretical ions for the listed peptide. Currently, the most often applied criteria in protein study are cross-correlation score between the observed peptide fragment mass spectrum and the theoretically predicted one (X_{corr}) and cross-correlation functions for the first and second ranked results (ΔC_n). Washburn et al. [7] applied the following criteria of correctness of peptide identification: X_{corr} above 1.9 for single charged fully tryptic peptides, over 2.2 and 3.75 for fully or partially tryptic doubly and triply charged peptides, respectively, and the ΔC_n values higher than 0.08. On the other hand, in the studies performed by Peng et al. [8] the peptides were classified as properly identified when X_{corr} was, in case of fully tryptic peptides, higher than 2.0, 1.5, or 3.3 for the charge states of 1+, 2+, 3+, correspondingly, and over 3.0 (2+ charged) or 4.0 (3+ charged) considering partially tryptic peptides, when ΔC_n score was above 0.08. The relationship between application of different filtering criteria and degree of false positive identifications has also been recently demonstrated by Qian et al. [9]. There it was shown that all previously applied filtering criteria were derived using either relatively simple proteomes (e.g., the yeast proteome) or standard proteins. The degree of false positive identifications, when these criteria are extended to considerably more complex mammalian proteomes, especially human proteome, is still problematic and requires improvement of the strategies to distinguish correct from incorrect ones. Therefore, to decrease the probability of random match, which is growing up with the size of the protein database, two new sets of filtering criteria were independently developed for human cell line and human plasma samples [9]. For human cell line samples, the new criteria were as follows: $X_{corr} \geq 1.5$ for fully tryptic peptides and $X_{corr} \geq 3.1$ for partially tryptic peptides for the 1+ charge state, $X_{corr} \geq 1.9$ for fully tryptic peptides and $X_{corr} \geq 3.8$ for partially tryptic peptides for 2+ charge state, and $X_{corr} \geq 2.9$ for fully tryptic peptides and $X_{corr} \geq 4.5$ for partially tryptic peptides for the 3+ charge state. All the criteria had ΔC_n value of ≥ 0.1. The new criteria for peptides from human plasma samples include for the 1+ charged, $X_{corr} \geq 2.0$ and ≥ 3.0 for fully and partially tryptic peptides, respectively; for the 2+ charged, $X_{corr} \geq 2.4$ for fully and ≥ 3.5 for partially tryptic peptides, consequently; and

for the 3+ charged, $X_{corr} \geq 3.7$ for fully and ≥ 4.5 for partially tryptic peptides, accordingly. The ΔC_n values were in all cases ≥ 0.1 as well.

Nevertheless, considering the variety and dynamic range of the proteins, occurring in the different organisms, there is still a possibility of false positive or false negative identification. Growing concerns about the quality of MS data affected in various ideas to harden protein identification by using bioinformatics' methods, for example, decoy search strategies [10] or additional information obtained during analysis, for example, peptide pI or retention time [11]. The retention time is very practical parameter in proteomics as it is easy to obtain from LC-MS data and does not require a lot of instrumental effort [2, 12]. Comparison of the experimental and predicted retention times of the occurring peptides may examine the correctness of the identification and then enable to exclude the incorrectly identified ones. However, to predict properly peptides' retention highly accurate models should be developed. Recently, some models have been proposed which characterize quantitatively the structure of a peptide and predict its gradient RP-LC retention at given separation conditions [13, 14].

Liquid chromatography (LC) is an analytical technique which can provide a great amount of quantitative, comparable, and reproducible (retention) data for large sets of structurally diversified compounds (analytes). On the other hand, chromatographic retention time can be considered as a chemical structure dependent parameter, which is constant for given separation conditions (mobile phase composition, stationary phase, temperature, pH). Due to that, quantitative structure (chromatographic) retention relationships (QSRR) have been considered as a model approach to establish strategy of retention predictions. However, to predict properly peptides' retention highly accurate models should be developed [15–17]. In particular, in proteomics, the structural descriptors obtained from QSRR studies can contribute to better predictions of retention times and therefore harden peptides identification.

Several previous reports [18–21] prove that retention of peptides in reversed-phase liquid chromatography (RP-LC) depends on their amino acids composition. There, the regression analysis was used to derive the regression coefficients, which represented the contribution of each amino acid in the peptide's sequence to its retention. This approach was applied in proteomics analysis, to predict the retention times of peptides' tryptic digests [22]. Then, it was also employed to increase the reliability of the peptides identification to check the predictive capability of artificial neural networks (ANNs) by Petritis et al. [23] or by Shinoda et al. [24], where created ANN was then applied to predict the retention times of peptides from *Escherichia Coli* proteome. The correlation between amino acid composition and peptide's retention time was used as well to provide the identity information, given by the tandem mass spectrometry, of the peptides from *Drosophila melanogaster* proteome, to exclude the false positive identifications [25].

Recently, a QSRR model based on multiple linear regression has been proposed [26] to quantitatively characterize

the structure of a peptide and to predict its gradient RP-LC retention at established separation conditions. The logarithm of the sum of gradient retention times of the amino acids composing the individual peptide, log $\mathrm{Sum_{AA}}$, the logarithm of the peptide Van der Waals volume, log $\mathrm{VDW_{Vol}}$, and the logarithm of its calculated n-octanol-water partition coefficient, clog$P$, were employed [26–29].

The aim of the study was to derive the retention time prediction model and check its predictive capability based on quantitative structure-retention relationships (QSRRs). The newly modified QSRR model was derived with the use of set of peptides identified with the highest scores and originated from eight model proteins [13, 24, 30–32]. Therefore, no synthesized peptides with known amino acid sequences were used to derive and check the model [14, 31]. Moreover, descriptors applied in the new QSRR model were obtained in the new, facilitated from practical point of view, manner. Finally, its predictive ability was supported by further investigation with the use of a *Bacillus subtilis* proteome digest (not like previously just applying synthesized peptides with known amino acid sequences). To demonstrate that ability three sets of testing peptides received from proteins identified with different levels of confidence were used. Moreover, the additional attempts were performed to demonstrate the utility of QSRR approach as the additional constraint confirming the correctness of the peptides identifications.

2. Material and Methods

2.1. Standards. The standard amino acids solutions were prepared by dissolving seven amino acids among twenty naturally occurring ones (isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine, and valine, all from Fluka BioChemika, Buchs, Switzerland) in 0.1% aqueous solution of trifluoroacetic acid (TFA). Water was deionized by passing through a Direct-Q (Millipore) system (Millipore, Bedford, MA, USA). The concentrations of the samples were approximately 0.6 mg/mL.

The solutions of standard proteins annotated as eight model proteins (about 3 mg/mL) were as follows: bovine serum albumin (BSA), chicken egg ovalbumin (CEO), bovine milk lactoglobulin (BML), bovine milk β-casein (BMC), bovine myoglobin (BM), human serum albumin (HSA) and ribonuclease B (RibB) from Sigma-Aldrich (Steinheim, Germany), and insulin-like growth factor-binding protein 1 (IGFBP-1), which was purified from human amniotic fluid following a previously reported procedure [33]. They were obtained by dissolving the lyophilized standard proteins in deionized water and then treated as shown below in digestion protocol.

2.2. Bacillus subtilis Sample Preparation

2.2.1. Growth Conditions. *Bacillus subtilis* strains were grown in nutrient broth (NB) supplemented with 0.2% KCl, 0.05% MgSO$_4$ (final concentration) and antibiotics, if appropriate with shaking at 37°C.

2.2.2. Spore Purification. As described before [33] forty-eight-hour cultures in nutrient broth were pelleted (10000 ×g, 10 minutes) and washed three times with 1/4 volume of cold water. The pellet was resuspended in 1/5 of the initial volume of cold MQ water and incubated overnight at 4°C. On subsequent days the suspension was centrifuged (20000 ×g, 20 min, 4°C). The pellet was resuspended in fresh cold MQ water. This procedure was repeated for 5 to 10 days. Purified spores were kept in water suspension at 4°C in the dark. Once per week the spore were centrifuged and suspended in fresh water to avoid spontaneous germination.

2.2.3. Protein Extraction. The spore pellet (approximately 20 mg spores) was resuspended in 1 mL of extraction buffer (50 mM Tris-HCl, pH = 7.8; 2% SDS; 10% glycerol; 0,2 M DTT) and boiled for 5 min and vortexed for 30 seconds. These steps were repeated twice. Unlysed spores and spore debris were removed by centrifugation at 12,000 ×g for 5 min at 4°C. The supernatant was precipitated with acidified acetone/methanol mixture. To one volume of protein solution four volumes of cold precipitation reagent were added and kept on at −20°C. Precipitate was spun down at 15,000 × g, at 4°C and supernatant was discharged an samples were drained, then resuspended in water, and stored at −80°C. Concentration of proteins was determined with the use of Bradford assay kit (Bio-Rad Laboratories) and it equalled 1.2–1.5 mg/mL.

2.3. Digestion Protocol. To 1 mL of each protein (BSA, CEO, BML, BMC, BM, HAS, RibB, and IGFBP-1) sample (∼3 mg/mL), 300 μL of DTT (dithiothreitol) (Sigma-Aldrich, Steinheim, Germany) (100 mM, freshly prepared in 100 mM ammonium bicarbonate buffer, pH 8.5) were added. The samples were kept in 60°C for 30 min, to allow reduction of the disulfide bridges. Then 50 μg of trypsin was added (ratio 1 : 50 E/S) to each sample. Samples were digested for 12 hours (overnight digestion) at 37°C. After that 0.1 mL of TFA was added to each sample to stop the digestion. Obtained standard solutions concentrations were about 50 pmol/μL.

To 1 mL of Bacillus subtilis spore cells lizates (1.2–1.5 mg/mL), 150 μL of DTT (Sigma-Aldrich, Steinheim, Germany) (100 mM, freshly prepared in 100 mM ammonium bicarbonate buffer, pH 8.5) were added. The samples were kept in 60°C for 30 min, to allow reduction of the disulfide bridges. Then 25 μg of trypsin was added (ratio 1 : 50 E/S) to each sample. Samples were digested for 12 hours (overnight digestion) at 37°C. After that 0.05 mL of TFA was added to each sample to stop the digestion. Obtained standard solutions concentrations were about 50 pmol/μL.

Tryptic digests were stored at −20°C (if frozen in this reaction mixture the disulfide bonds would not reoxidase). The LC-ESI-MS/MS analyses were performed in three weeks at the latest (the shelf life of such frozen solution is couple of months) (http://www.thermo.com/).

2.4. LC Conditions. The chromatographic analysis was performed on C-18 analytical column: XTerra MS C18 3.5 μm (2.1 × 100 mm) column (Waters, Milford, MA, USA).

The mobile phase consisted of two solvents (A and B) mixed on-line. Solvent A was 0.1% aqueous (water was MS-grade) solution of trifluoroacetic acid (TFA) (Sigma-Aldrich, Steinheim, Germany) and solvent B was acetonitrile (ACN) (MS-grade, Sigma-Aldrich, Steinheim, Germany) containing 0.1% TFA. The applied linear gradient time was 90 min, from 0% B to 60% B. The flow rate was 200 μL/min. The injection volume was 10 μL. The LC-MS apparatus was equipped with thermostated column oven and surveyor autosampler controlled at 20°C (Thermo Finnigan, San Jose, CA, USA), a quaternary gradient Surveyor MS pump (Thermo Finnigan, San Jose, CA, USA) with a diode array detection (DAD) system, and LTQ linear ion trap MS system with ESI ion source controlled by Xcalibur software 1.4 (Thermo Finnigan, San Jose, CA, USA).

2.5. MS Conditions. The MS/MS analysis was performed on Finnigan LTQ instrument (Thermo Finnigan, San Jose, CA, USA). Mass spectra were generated in positive ion mode under constant instrumental conditions: source voltage 4.62 kV, capillary voltage 40.97 V, sheath gas flow rate 39.99 (arbitrary units), auxiliary gas flow 10 (arbitrary units), sweep gas flow 0.95 (arbitrary units), capillary temperature 219.96°C, and tube lens voltage 250.43 V. MS/MS spectra, obtained by CID (collision-induced dissociation) in the linear ion trap, were performed with an isolation width 3Da (m/z); the activation amplitude was 35% of ejection RF amplitude that corresponds to 1.58 V.

2.6. Protein Identification. The experimental retention times of the peptides ($t_{R\,\mathrm{exp}}$) were determined at peak intensity maximum. The m/z values measured manually for the most intense peaks in acquired MS/MS spectra were automatically searched against the protein database (*fasta) using the Sequest Algorithm, incorporated into Bioworks 3.0 (Thermo Finningan, San Jose, CA, USA). The *fasta format for each protein was downloaded from Expasy (http://www.expasy.org/sprot/). During the interpretation of the results obtained after the correlation analysis done on the experimental and the predicted retention times of peptides, the exemplary filtering criteria applied in the studies were the same as those discussed previously, proposed by Washburn et al. [7]. The spectra for singly charged peptides with a cross-correlation score to a tryptic peptide (X_{corrs}) greater than 1.9, the spectra for doubly charged tryptic peptides with X_{corrs} of at least 2.2, and the spectra for triply charged tryptic peptides with X_{corrs} above 3.75 were accepted as correctly identified according to Sequest software. For all the spectra analyzed, ΔC_n values were above 0.08.

2.7. QSRR Analysis. Multiple regression equations for model set of peptides based on the experimental retention times were derived by employing Microsoft Excel software (Microsoft Co., Redmond, WA, USA) and Statistica (StatSoft, Tulsa, OK, USA) run on a personal computer.

Regression coefficients (± standard deviations), multiple correlation coefficients, R, standard errors of estimate, s, significance levels of each term and of the whole equations, p, and values of the F-test of significance, F, were calculated.

The structural descriptors of the analyzed standard amino acids and peptides from investigated, standard proteins and *Bacillus subtilis* cells were calculated. First of all, in contrary to the previous models [26–29], where just log $\mathrm{Sum_{AA}}$ was calculated by simple addition of component amino acids retention (taking into account all 20 naturally occurring amino acids), the novel QSRR peptide descriptor log Sum $(k + 1)_{\mathrm{AA}}$ was used. The retention factor (k) was introduced, because it is more similar for different related systems than t_R as it compensates for some physical differences between columns. Descriptor log Sum $(k + 1)_{\mathrm{AA}}$ was calculated applying retention data for just only 7, the most retained amino acids (isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine, and valine). The other 13 amino acids are hardly retained; therefore their presence in peptide's sequence does not influence significantly its retention. For these 13 amino acids fixed values were ascribed ($k = 0$) and one was added to avoid zero in the calculation of the logarithm, according to the procedure elaborated and evaluated elsewhere [34]. On the other hand, searching for the most accurate the logarithm of its calculated n-octanol-water partition coefficient, clog P, values, different calculation methods were tested (data not shown). Briefly, to obtain clog P values HyperChem 7.5 professional software for personal computers (HyperCube, Waterloo, Canada) with the extension ChemPlus, Dragon professional 5.0 software (Milano Chemometrics and QSAR Research Group—Talete, Milano, Italy), and on-line available ALOGPS 2.1 software (http://www.vcclab.org/) were obtained. Finally, to derive the appropriate QSRR model, clog P values average, log P module in ALOGPS 2.1 software was used to determine that QSRR descriptor.

The general QSRR equation has the following form:

$$t_R = k_1 + k_2 \log \mathrm{Sum}\ (k + 1)_{\mathrm{AA}} + k_3 \mathrm{clog}\ P, \qquad (1)$$

where t_R is the gradient HPLC retention time and k_1-k_4 are regression coefficients.

3. Results and Discussion

3.1. Derivation and Validation of QSRR Model. The QSSR model was derived from peptides obtained from the digestion of 8 model proteins. The amino acid sequences of these peptides were proved by MS/MS analysis and identified by Sequest software (Bioworks 3.0 package Thermo Fisher Scientific Inc., Waltham, MA, USA). Only peptides with the highest scores were taken into account in the model set of peptides used to derive the QSRR model. Peptides were assumed and considered as true positives according to their cross-correlation score to a tryptic peptide X_{corr} values with over 2.0 for 1+ and 2+ and over 4.5 for 3+ charged peptides. Peptides with lower values of X_{corr} were excluded from the model set of peptides, due to higher possibility of their false positive identification. Hence, the peptides included in

the study were divided into five groups: one set of model peptides (Table 1) and four testing sets of peptides (Tables 2–5). 50 model peptides used to derive QSRR model and collected in Table 1 originated from 8 model proteins. The 21 peptides reported in Table 2 were used to check the general validity of the proposed QSRR model. In view of the main objective of this work, three other sets of testing peptides originating from *B. subtilis* proteome digestion were used. One set includes 54 peptides belonging to proteins identified on the basis of more than one peptide with X_{corr} above 1.5 (Table 3). A second set comprises 41 peptides belonging to proteins identified again with X_{corr} above 1.5, but on the basis of just one peptide (Table 4). And the third set comprises 40 peptides belonging to proteins identified on the basis of just one peptide, but with X_{corr} below 1.5 (Table 5).

The model set consisting of 50 peptides with the highest values of X_{corr} was used to create a model to predict further retention times of the peptides from proteome of *Bacillus subtilis* cells. Among this group differences between experimental and predicted retention times ranged from 0.01 to 2.81 min. 42% (21 peptides) of the results were characterized by differences between experimental and predicted retention times lower than 1 min, and for the remaining 58% (29 peptides), these values ranged from 1 to 3 min (Table 2). Taking into account retention times and the values of descriptors for those 50 model peptides, the following specific equation was derived:

$$t_R = -17.53\ (\pm\ 1.54) + 32.18\ (\pm\ 1.10)\log\ \text{Sum}\ (k+1)_{AA}$$
$$+\ 0.76\ (\pm\ 0.10)\text{clog}\ P,$$
$$p = 4 \times 10^{-15}, \quad p = 9 \times 10^{-32}, \quad p = 7 \times 10^{-10},$$
$$\text{with } n = 50, \quad R = 0.974, \quad s = 1.45, \quad F = 431,$$
$$p < 6 \times 10^{-31}.$$

$$(2)$$

The description of t_R by (2) was good as documented by the following criteria of statistical quality. All the regression coefficients were highly statistically significant as was the whole equation. Multiple correlation coefficient, R, standard error of estimate, s, and the value of the F-test of significance, F, all were also satisfactory.

Equation (2) provides the predictive model based on experimentally obtained descriptor $(\log\ \text{Sum}\ (k+1)_{AA})$ and improved by the implementation of molecular-modeling-based descriptor $(\text{clog}\ P)$. Experimentally obtained descriptor $(\log\ \text{Sum}\ (k+1)_{AA})$ appeared to possess significant contributions into peptides' retention. However, the $\log\ \text{Sum}\ (k+1)_{AA}$ has little in common with *n*-octanol/water partition coefficient—neither for individual amino acids nor for the peptide. The considered analytes were highly ionizable and only minute fraction of molecules can exist in nonionized form in solution. Only for that fraction $\log\ P$ $(\text{clog}\ P)$ properly reflects the ability to partition between aqueous and hydrophobic phase. Therefore, the $\log\ \text{Sum}\ (k+1)_{AA}$ parameter was not considered to mimic $\text{clog}\ P$; actually it reflects differences in
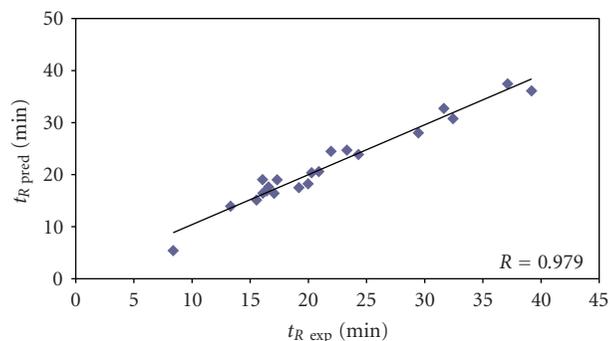
FIGURE 1: Correlation between experimental and predicted retention times for a set of test peptides obtained from model proteins ($n = 21$).

peptides polarities. Instead, $\text{clog}\ P$ was an auxiliary peptide structure descriptor: a correction for $\log\ \text{Sum}\ (k+1)_{AA}$.

In order to check the correctness of the model, the set of 21 peptides (Table 2), derived from 8 model proteins, was used as the validation set. The predicted retention times, calculated from (2), were then compared to the experimental retention times and the differences between these two retention times were calculated. Differences varied from 0.09 to 3.08 minutes in retention time (mean value 1.29 min, Table 2). For 9 peptides the range of differences between experimental and predicted retention times (42.86%) was from 0.09 to 0.46 min; for 11 peptides (52.38%) the range was 1.07–2.99 min; for 1 peptide (4.76%) this value was over 3 min. Correlation ($R = 0.979$) between experimental and predicted retention times confirmed additionally the validity of the model (Figure 1), proving that similar values of predicted and experimental retention times of analyzed peptides correlate also with higher probability of identification correctness using Sequest algorithm (Figure 5).

3.2. QSRR-Based Analysis of Peptides from Bacillus subtilis Proteome. Using (1), the predicted retention times for peptides identified for proteome of *Bacillus subtilis* cells were further calculated (Tables 3–5). The experimental retention times for these peptides were obtained in LC-MS/MS analysis and compared to the calculated ones. Here, the special attention on peptides with low X_{corr} (around 1.5) was taken into account to check the applicability of the proposed model and to indicate the potential false positives. In this case, the most important were the attempts to provide the QSRR-based tool to confirm true and false positively identified peptides.

The derived accurate model, as confirmed in Figure 1, was applied to calculate also the retention times of peptides from the real proteome sample of *Bacillus subtilis* cells. Its correctness was proved first by calculating the predicted retention times of peptides belonging to proteins identified on the basis of more than one peptide with X_{corr} above 1.5, that is, those ones that are assumed to be the most confident true positives. It is clearly seen on correlation plot depicted in Figure 2 that the predicted retention times and

TABLE 1: Model peptides used to derive QSRR model.

Peptide sequence	Protein	m/z	Missed cleavages	Charge	X_{corr}	$\log \mathrm{Sum}_{(k+1)\mathrm{AA}}$	$c\log P$	$t_{R\ \exp}$
ALKALPMHIR	1	575.73	1	2	3.06	1.3542	−1.74	25.12
LFTFHADICTLPDTEKQIK	3	1111.28	1	2	3.21	1.6674	−4.6	32.60
KIKVYLPR	4	509.15	2	2	2.27	1.3005	−0.95	24.06
LVNEVTEFAK	6	575.65	0	2	3.42	1.3540	−2.44	24.53
YTRKVPQVSTPTLVEVSR	1	1031.19	2	2	3.01	1.4758	−5.67	25.80
ALHVTNIK	8	896.07	0	1	2.48	1.2148	−3.14	19.69
DTHKSEIAHR	3	597.64	1	2	2.8	1.1246	−7.77	13.15
AAFTECCQAADK	6	687.7	0	2	3.07	1.2657	−5.69	19.31
ALPGEQQPLHALTR	8	766.37	0	2	3.37	1.4145	−5.22	24.36
VKEAMAPK	5	874.08	1	1	2.13	1.0165	−2.19	14.40
QHMDSSTSAASSSNYCNQMMK	7	789.84	0	3	4.75	1.4546	−11.8	20.46
AFDEKLFTFHADICTLPDTEK	3	814.91	1	3	5.11	1.7127	−4.51	34.45
HIIVACEGNPYVPVHFDASV	7	1113.73	0	2	5.15	1.6128	−3.51	32.15
VHTECCHGDLLECADDRADLAK	6	1295.34	1	2	4.47	1.5447	−9.95	24.98
AFDEKLFTFHADICTLPDTEKQIK	3	1406.60	2	2	4.42	1.7629	−4.95	34.77
DLGEENFK	6	952	0	1	2.09	1.2654	−4.48	19.46
CCAAADPHECYAK	6	778.79	0	2	3.46	1.2195	−5.77	16.98
IPGSPEIR	8	869.00	0	1	2.18	1.1657	−2.43	19.40
LKPDPNTLCDEFKADEKKFWGKYLYEIAR	1	1174.00	5	3	4.49	1.8691	−5.45	39.20
ALHVTNIKK	8	1024.24	1	1	2.01	1.2405	−4.64	18.27
VLPVPQKAVPYPQR	5	796.96	1	2	3.32	1.3948	−3.24	24.20
AEFAEVSK	6	880.97	0	1	2.10	1.1909	−2.94	18.06
ALHVTNIKK	8	512.62	1	2	2.70	1.2405	−4.64	18.24
TCVADESAENCDK	6	751.23	0	2	4.05	1.1488	−7.23	15.32
NECFLQHK	6	1077.16	0	1	2.33	1.2654	−3.74	19.67
TCVADESHAGCEK	3	675.73	0	2	3.47	1.1488	−7.76	14.99
CASIQKFGER	3	570.16	1	2	2.78	1.2958	−4.66	19.81
VHTECCHGDLLECADDR	6	1046.05	0	2	6.23	1.4161	−8.42	22.86
LFTFHADICTLPDTEK	3	926.55	0	2	4.79	1.6039	−3.25	32.90
RIPGSPEIR	8	513.09	1	2	2.63	1.1944	−3.5	20.00
HLVDEPQNLIK	3	653.75	0	2	3.48	1.3690	−4.68	24.56
TCVADESHAGCEKSLHTLFGDELCK	3	1348.00	1	2	3.77	1.6483	−7.37	31.14
YPNCAYK	7	916.99	0	1	2.58	1.1508	−2.34	16.78
LRCASIQKFGER	1	704.83	2	2	3.82	1.4107	−4.96	22.79
WKEPCRIELYR	8	747.38	2	2	2.70	1.4991	−3.52	26.67
TPEVDDEALEKFDK	1	818.87	1	2	5.09	1.4067	−6.12	24.94
LDELRDEGK	6	538.08	1	2	2.52	1.2299	−5.34	16.62
YICDNQDTISSKLK	1	814.91	1	2	3.80	1.4504	−6.1	22.75
YICDNQDTISSKLK	3	814.91	1	2	4.03	1.4504	−6.1	22.75
QTALVELLKHKPK	3	753.42	2	2	2.89	1.4159	−3.08	27.89
VKEAMAPKHK	5	570.20	2	2	2.90	1.0930	−3.41	13.48
ELINSWVESQTNGIIR	4	930.53	0	2	3.93	1.6097	−5.86	31.95
ISQAVHAAHAEINEAGR	4	887.96	0	2	3.19	1.3932	−7.56	19.51
YICDNQDTISSK	3	694.25	0	2	3.8	1.3468	−6.61	18.79
SHCIAEVEKDAIPENLPPLTADFAEDKDVCK	1	1133.93	2	3	5.65	1.7342	−3.53	33.13
GGLEPINFQTAADQAR	4	844.91	0	2	4.19	1.4735	−6.71	27.41
EAMAPKHKEMPFPK	5	821.49	2	2	3.87	1.3625	−3.31	21.52
FYLPNCNKNGFYHSR	8	931.04	0	2	2.76	1.5912	−5.83	26.57
YICENQDSISSK	6	723.25	0	2	4.04	1.3468	−6.77	18.06
SLHTLFGDELCK	3	682.28	0	2	3.55	1.4829	−3.32	30.69

TABLE 2: Test peptides obtained from a set of model proteins and used to check the validity of the proposed QSRR model.

Peptide sequence	Protein	m/z	Missed cleavages	Charge	X_{corr}	$\log \text{Sum}_{(k+1)AA}$	$c\log P$	$t_{R\,exp}$	$t_{R\,pred}$	Dt_R
WKEPCR	8	818.97	1	1	1.50	1.1950	−2.52	17.32	19.00	1.69
VVESLAK	8	745.89	0	1	1.52	1.1192	−2.11	16.42	16.88	0.46
VLPVPQK	5	780.98	0	1	1.57	1.1192	−1.30	19.19	17.49	1.70
IELYR	8	693.81	0	1	1.57	1.2011	−0.67	20.90	20.61	0.29
LDELR	6	645.73	0	1	1.62	1.1133	−2.48	17.06	16.40	0.66
RIPGSPEIR	8	1025.19	1	1	1.68	1.1944	−3.50	19.97	18.24	1.74
NGFYHSR	8	880.93	0	1	1.72	1.2308	−3.98	16.06	19.04	2.99
SLGKVGTR	3	817.96	1	1	1.79	1.1164	−4.29	15.54	15.13	0.41
AQETSGEEISK	8	1179.22	0	1	1.83	1.1560	−7.53	13.32	13.94	0.62
NVACK	7	592.65	0	1	1.90	0.7843	−3.04	8.38	5.39	2.99
ETCFAEEGKK	6	600.63	1	2	2.00	1.2158	−5.18	16.58	17.65	1.07
CCAADDKEACFAVEGPKLVVSTQTALA	1	1371.57	2	2	2.01	1.6500	−6.31	32.44	30.76	1.68
FYLPNCNK	8	500.08	0	2	2.02	1.3425	−2.34	24.31	23.88	0.43
HLKTEAEMK	2	544.14	1	2	2.12	1.1551	−4.21	16.09	16.43	0.34
HKEMPFPK	5	507.61	1	2	2.25	1.1970	−0.82	20.27	20.36	0.09
ALKAWSVAR	3	501.60	1	2	2.28	1.3755	−2.65	23.31	24.71	1.40
LFTFHADICTLPDTEKQIKK	1	783.91	2	3	2.57	1.6767	−4.86	31.65	32.72	1.07
TPEVDDEALEKFDKALK	1	650.38	2	3	2.58	1.5119	−4.06	29.46	28.03	1.43
LYAEERYPILPEYLQCVKELYR	4	930.74	2	3	2.59	1.7534	−3.66	39.18	36.10	3.08
LFTFHADICTLPDTEKQIKKQTALVELLK	1	1115.98	3	3	3.22	1.8423	−5.65	37.13	37.45	0.32
LKECCDKPLLEK	1	710.37	2	2	4.32	1.3797	−3.12	21.94	24.49	2.55

experimental retention times do not vary significantly, and so it can be concluded that those peptides, and the proteins, to which they are assigned, are correctly identified and really present in the analyzed sample. The detailed accuracy of the peptide identification can be further examined in Table 3. In the set of 54 peptides obtained from digestion of *Bacillus subtilis* proteome and belonging to proteins identified on the basis of more than one peptide with X_{corr} above 1.5, the differences between experimental and predicted retention times varied from 0.08 to 18.07 min (mean value 5.13 min). For 8 peptides, being 14.82% of the set, the difference between experimental and predicted retention times was lower than 1 min. There were 6 peptides (11.11%), which retention times differences ranged between 1 and 3 min. In most cases, differences between experimental and predicted retention times were from 3 to 5 min and then from 5 to 10 min, for 18 (33.33%) and 16 (29.63%) peptides, respectively. 4 peptides (7.41%) were characterized by difference in experimental and predicted retention times ranging from 10 to 15 min. There were even also 2 cases, for which these values varied between 15 and 20 min. The correlation between experimental and predicted retention times can be considered good with correlation coefficient equaled 0.936 (Figure 2). However, some peptides in this set could be considered probably as false positives (e.g., ESIAQVAAISAADEEVGSLIAEAMER, or MSGWLAHILE-QYDNNRLIRPR). Generally, at that moment, it was proved that it is again possible to predict the retention times of unknown peptides of *Bacillus subtilis* proteome, based on retention data obtained experimentally only for the limited
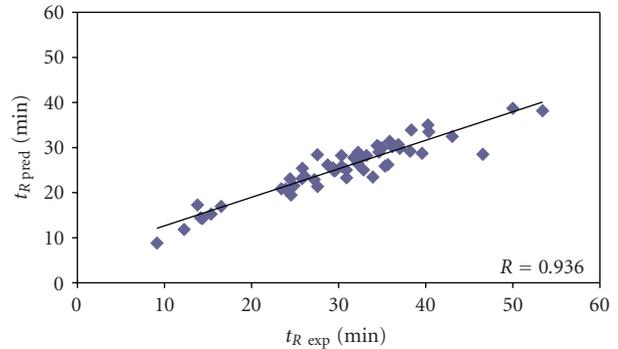

FIGURE 2: Correlation between experimental and predicted retention times for a set of test peptides obtained from *Bacillus subtilis* proteome. The proteins were identified on the basis of more than one peptide with X_{corr} above 1.5 ($n = 54$).

number of known model peptides originating from 8 known model proteins.

Among 41 *Bacillus subtilis* peptides belonging to proteins identified on the basis of only just one peptide with X_{corr} above 1.5 (Table 4), the difference between experimental and predicted retention times varied from 0.35 to 11.7 min and the mean value was 4.92 min. The predicted retention times of 5 peptides varied from the experimental ones less than 1 min, which refers to 12.20% of the investigated set. For other 8 peptides (19.51%) the difference between experimental and predicted retention

TABLE 3: Test peptides from proteins of Bacillus subtilis proteome, identified on the basis of more than one peptide with X_{corr} above 1.5.

Peptide sequence	m/z	Missed cleavages	Charge	X_{corr}	$\log \mathrm{Sum}_{(k+1)AA}$	clog P	$t_{R\,\text{exp}}$	$t_{R\,\text{pred}}$	Dt_R
ALDMLEASPVQGFDAK	846.96	0	2	4.90	1.52	−4.71	32.56	27.66	4.90
ITGTSNYEDTAGSDIVVITAGIAR	1213.32	0	2	4.87	1.63	−6.38	34.61	30.19	4.42
RHDDYDSKK	582.61	1	2	2.57	1.10	−7.90	12.26	11.84	0.42
KPHHHCDDYK	640.70	1	2	3.17	1.13	−6.12	14.34	14.26	0.08
DYLYQEPHGK	625.67	0	2	2.60	1.33	−4.72	24.86	21.52	3.34
EGLKDYLYQEPHGK	839.42	1	2	2.81	1.46	−5.99	30.89	25.03	5.86
KEGLKDYLYQEPHGK	903.51	2	2	3.71	1.48	−4.79	32.18	26.41	5.77
YYKKPHHHCDDYK	867.96	2	2	1.94	1.38	−4.29	33.93	23.46	10.47
GTAMAYDQIDGAPEER	862.92	0	2	3.63	1.38	−8.06	23.43	20.86	2.57
TVGSGVVSTITE	1150.26	0	1	1.42	1.27	−5.07	24.54	19.44	5.10
GITISTAHVEYETETR	904.47	0	2	4.86	1.44	−7.39	24.43	23.06	1.37
GQVLAKPGTITPHSK	767.89	1	2	2.77	1.37	−6.82	27.57	21.33	6.24
VGDEVEIIGLQEENKK	900.99	1	2	4.59	1.46	−6.08	29.49	24.80	4.69
HYAHVDCPGHADYVK	856.94	0	2	4.62	1.39	−4.95	30.91	23.30	7.61
DLLSEYDFPGDDVPVVK	955.04	0	2	4.92	1.58	−4.46	35.03	30.00	5.03
LLDYAEAGDNIGALLR	852.95	0	2	3.63	1.59	−4.49	36.17	30.20	5.97
NMITGAAQMDGAILVVSAADGPMPQTR	1359.07	0	2	5.04	1.64	−7.20	37.02	29.78	7.24
SHANIGTIGHVDHGKTTLTAAITTVLHKKSGK	1099.25	3	3	3.27	1.72	−11.83	39.57	28.71	10.86
NVGVPYIVVFLNKCDMVDDEELLELVEMEVR	1806.60	1	2	4.26	1.85	−4.97	53.42	38.15	15.27
ALAPEIVGEEHYAVAR	863.46	0	2	2.54	1.46	−4.87	30.36	25.85	4.51
EGNDLFYEMSDSGVINK	960.03	0	2	4.61	1.56	−6.57	31.77	27.82	3.95
GMEAVDTGAPISVPVGDVTLGR	1071.71	0	2	5.50	1.54	−5.68	32.56	27.81	4.75
VFNVLGENIDLNEPVPADAK	1078.20	0	2	5.91	1.61	−5.15	34.41	30.44	3.97
KLTEMGIYPAVDPLASTSR	1025.68	1	2	3.85	1.56	−4.80	34.63	29.00	5.63
VQPGQQHLKR	596.18	1	2	2.37	1.18	−6.77	15.36	15.23	0.13
IVSINPADKEEVVGR	813.92	0	2	3.32	1.40	−5.93	27.21	22.88	4.33
AGGPDYLALHMQAK	736.85	0	2	3.68	1.43	−4.63	29.65	24.93	4.72
VSDFDEALEVANNTEYGLTGAVITNNRK	1014.75	1	3	4.37	1.72	−9.43	36.85	30.66	6.19
GYFIKPTIFADLDPK	863.50	1	2	3.55	1.62	−1.12	38.36	33.88	4.48
LMQEEIFGPVVAFCK	856.54	0	2	3.25	1.59	−0.08	40.34	33.48	6.86
QQNQSAEQNKQQNS	816.82	1	2	2.81	1.15	−13.83	9.13	8.82	0.31
KQNQQSAAGQGQFGTEFASETNAQQVR	1456.52	1	2	5.75	1.61	−11.78	25.84	25.38	0.46
YDDYDKK	946.98	1	1	1.90	1.15	−2.97	13.79	17.24	3.45
DYDCDYDKK	583.11	1	2	2.76	1.21	−5.81	16.52	16.93	0.41
DYDYVVEYK	597.64	0	2	3.30	1.34	−3.31	25.78	23.08	2.70
DYDYVVEYKK	661.72	1	2	2.57	1.36	−3.32	26.07	23.70	2.37
VGNDGVITIEESK	681.24	0	2	2.66	1.34	−6.20	23.92	20.83	3.09
FGSPLITNDGVTIAK	767.38	0	2	2.73	1.52	−4.16	30.33	28.23	2.10
EIELEDAFENMGAK	798.87	0	2	3.31	1.46	−5.96	32.86	25.01	7.85
ESIAQVAAISAADEEVGSLIAEAMER	1331.46	0	2	6.41	1.64	−8.79	46.55	28.48	18.07
WNTNAGDDYVSNGPFK	893.43	0	2	2.63	1.57	−5.83	27.55	28.40	0.85
GVIMPGTGEVYFR	713.83	0	2	2.42	1.47	−1.26	32.23	28.95	3.28
ADYTGPDKQK	562.10	1	2	2.85	1.13	−5.89	14.13	14.44	0.31
MLTEIGEVENAEPYIR	933.05	0	2	4.38	1.51	−4.61	31.8	27.65	4.15
EDYGIAENFLYTLNGEEPSPIEVEAFNK	1595.71	0	2	3.18	1.80	−7.22	40.25	35.03	5.22
MSGWLAHILEQYDNNRLIRPR	861.99	2	3	3.54	1.73	−7.62	43.05	32.45	10.60
VLQQPNCLEVTISPNGNK	978.11	0	2	4.43	1.50	−5.93	28.75	26.15	2.60
YRDNNYLDDEHEVIAK	998.05	1	2	4.61	1.50	−6.85	29.35	25.49	3.86

TABLE 3: Continued.

Peptide sequence	m/z	Missed cleavages	Charge	X_{corr}	$\log \mathrm{Sum}_{(k+1)AA}$	clog P	$t_{R\ exp}$	$t_{R\ pred}$	Dt_R
IVVQAEREFLAEVVGETK	1009.65	1	2	2.99	1.56	−4.48	38.2	29.17	9.03
IVNPLGQPVDGLGPILTSK	960.13	0	2	5.18	1.60	−3.36	35.86	31.39	4.47
KGRNPQTGEEIEIPASKVPAFKPGK	894.02	4	3	4.53	1.59	−7.21	33.22	28.24	4.98
MNKTELINAVAEASELSK	975.11	1	2	5.11	1.50	−6.38	35.32	25.91	9.41
MNKTELINAVAEASELSKK	1039.19	2	2	5.25	1.51	−6.58	35.64	26.19	9.45
AVDSVFDTILDALKNGDKIQLIGFGNFEVR	1099.57	2	3	5.81	1.87	−5.30	50.01	38.68	11.33

TABLE 4: Test peptides from proteins of *Bacillus subtilis* proteome, identified on the basis of one peptide with X_{corr} above 1.5.

Peptide sequence	m/z	Missed cleavages	Charge	X_{corr}	$\log \mathrm{Sum}_{(k+1)AA}$	clog P	$t_{R\ exp}$	$t_{R\ pred}$	Dt_R
NIAEMVK	754.79	0	1	1.62	1.1042	−2.43	14.25	16.15	1.90
INIMSAR	1463.68	0	1	1.87	1.1746	−2.86	14.23	18.09	3.86
NLLFAAR	707.86	0	1	1.81	1.3307	−1.12	14.27	24.44	10.17
LNSLDSR	896.11	0	1	1.64	1.1755	−4.79	14.29	16.65	2.36
DIMSPSR	1380.53	0	1	1.53	1.0654	−4.27	14.30	13.50	0.80
LALDLESKK	922.17	1	1	1.62	1.3216	−2.73	18.01	22.92	4.91
IDIALESKK	1020.08	1	1	1.54	1.2931	−2.77	18.04	21.97	3.93
SHTGKAAVLNR	524.07	1	1	1.52	1.2061	−6.85	24.54	16.07	8.47
GHNPGQPEPLSGSK	718.86	0	2	3.54	1.2550	−8.65	17.05	16.27	0.78
VVVSVNTDQDQAQAQSQDGED	868.09	0	2	4.73	1.4038	−14.74	19.37	16.42	2.95
GNQVSENLQQAAR	694.78	0	2	2.03	1.2571	−9.4	20.52	15.76	4.76
LIDKHKKYVYHRINK	920.72	4	2	2.60	1.5299	−4.29	29.00	28.43	0.57
EAEELIPNVTTAAVK	1025.52	0	2	2.43	1.3889	−6.37	29.18	22.31	6.87
ELQEKFLIPAVEQKK	1044.81	2	2	2.24	1.5292	−3.62	29.27	28.92	0.35
QDIPIEARMNEIVHSLK	1098.25	1	2	2.15	1.5231	−5.67	29.33	27.16	2.17
AAEMAVARQNEQKVKK	617.20	3	2	2.20	1.2894	−7.16	29.44	18.51	10.93
EGTVIKELIGAGQLDEK	817.41	1	2	2.40	1.5147	−5.74	29.49	26.84	2.65
EVMIEGVLSVLEGQAPK	731.84	0	2	2.38	1.5167	−4.34	29.51	27.97	1.54
DRVFIAPVGGGPR	580.16	1	2	2.68	1.3967	−3.64	29.82	24.64	5.18
SGETEDSTIADIAVATNAGQIK	865.45	0	2	3.33	1.5192	−9.39	30.05	24.21	5.84
IDNLSYYIEQEYK	952.72	0	2	2.13	1.5361	−4.63	30.89	28.37	2.52
SGSIESIDVSLTDLR	613.73	0	2	2.53	1.4873	−6.49	33.05	25.39	7.66
LEIASEFGVNLGADTTSR	1481.93	0	2	4.30	1.5661	−5.87	33.12	28.39	4.73
HSSDEEPFSALAFK	531.67	0	2	2.95	1.4894	−5.34	33.16	26.33	6.83
AVLSPLFPTATEGGENMDSNLK	1146.78	0	2	4.62	1.6314	−6.36	34.13	30.12	4.01
VCELQKVAVLNINDLANAVK	1078.27	1	2	2.00	1.5981	−4.6	34.35	30.39	3.96
TEWRQERLNPLQRLTGR	1077.71	3	2	2.48	1.5870	−8.41	34.47	27.13	7.34
GVSNNIIELINASGEPVIWK	1077.73	0	2	2.25	1.6913	−4.42	34.49	33.52	0.97
LSLKSIIGGRIPNYHK	955.65	2	2	2.06	1.6217	−5.35	35.03	30.58	4.45
ANVPLDQIAVLSIGTGEAPTR	1062.20	0	2	4.11	1.5774	−5.59	35.54	28.97	6.57
DQDISGEKATADQLLKDVK	1038.13	2	2	2.09	1.4968	−8.56	35.61	24.12	11.49
LIDIVNPTPQTVDALMR	949.11	0	2	4.64	1.5453	−4.51	36.18	28.76	7.42
AEELGAIIVDPSKTDDVVAEIAER	1271.40	1	2	2.49	1.6150	−6.83	36.41	29.24	7.17
GGGFLIEDVTYDQMYTPEDFTDEHK	1455.04	0	2	2.46	1.7382	−7.29	36.58	32.85	3.73
AIDSAVEELTFIAGQKPVVTR	1123.28	1	2	2.89	1.6162	−5.29	37.19	30.45	6.74
TYNLSLDNGGDFIQIGSDGGLLPR	1262.37	0	2	3.54	1.7529	−7.58	37.31	33.10	4.21
TIPLNITPYYASLMDPDNPR	1146.80	0	2	2.01	1.6343	−5.13	37.49	31.15	6.34
IVPISEIPSDLEAIDIGTK	1006.15	0	2	2.95	1.6095	−3.93	37.73	31.27	6.46
IQNGDPIAGLFDEFTQTVQR	1125.73	0	2	2.68	1.6493	−5.7	42.90	31.20	11.70
KVKTINRQIKISIRAEDQAFYR	893.71	5	3	2.54	1.6664	−7.19	33.22	30.62	2.60
SLEEGQEVSFEIVEGNRGPQASNVVKL	973.06	2	3	2.52	1.6909	−8.47	34.44	30.43	4.01

Table 5: Test peptides from proteins of *Bacillus subtilis* proteome, identified on the basis of one peptide with X_{corr} below 1.5.

Peptide sequence	m/z	Missed cleavages	Charge	X_{corr}	$\log Sum_{(k+1)AA}$	clog P	$t_{R\,exp}$	$t_{R\,pred}$	Dt_R
RADGSINQHPQER	754.79	1	2	1.4014	1.2128	−10.28	14.94	13.67	1.27
KGTDWNLYFWTAASYNAVIFVFVLV	1463.68	1	2	1.0082	1.9367	−5.79	42.39	40.38	2.01
ALECFKEMTTKI	707.86	2	2	1.0212	1.4322	−5.63	26.63	24.27	2.36
VKVIKPDP	896.11	2	1	0.9391	1.1301	−1.63	15.64	17.59	1.95
AQLSEKKGADGYL	1380.53	2	1	1.1544	1.3902	−5.26	26.46	23.20	3.26
TRLMGLILAVVAVGMIGAG	922.17	1	2	1.0633	1.6382	−6.61	33.61	30.15	3.46
SDNNIDKTL	1020.08	1	1	1.2258	1.2125	−8.54	18.87	14.98	3.89
EEKENWVL	524.07	1	2	0.9181	1.3569	−5.1	26.26	22.25	4.01
SWIGLPAPIFAGIAAIFAIQP	718.86	0	3	1.2819	1.8072	−3.13	33.64	38.24	4.60
LLGILTGFFMIGAKRP	868.09	2	2	0.9776	1.6883	−3.32	39.42	34.27	5.15
ELSASMG	694.78	0	1	1.1685	1.0896	−5.98	18.33	12.98	5.35
KHGVHIVAGSVAVRKNSDVYNTMYI	920.72	3	3	1.2386	1.6583	−11.81	33.61	26.84	6.77
DGWKVCLGKVGSMDAHKVVAAIETASKKSG	1025.52	5	3	1.2037	1.7212	−13.46	37.72	27.61	10.11
EYLDLLEKNVPYPAPSDLIFWSNEDY	1044.81	1	3	1.1465	1.8645	−4.77	48.33	38.83	9.50
KAEDLLRKVGLFEKRNDY	1098.25	5	2	1.0209	1.6135	−5.98	39.52	29.83	9.69
LLFKPNEERS	617.20	2	2	1.1289	1.3877	−3.86	10.37	24.18	13.81
EVTPEIEAAAGKGFTI	817.41	1	2	1.0159	1.4795	−9.51	9.03	22.84	13.81
NRVEYVKAEIQI	731.84	2	2	1.0301	1.3873	−5.22	40.09	23.13	16.96
LEEFKKDLH	580.16	2	2	0.8629	1.3695	−3.24	6.94	24.07	17.13
AGQHERLKEMNVTDT	865.45	2	2	1.0166	1.3300	−8.47	37.37	18.82	18.55
TGALIVYTSADSVLQIAAHEEVVPLEE	952.72	0	3	1.1387	1.7286	−5.69	52.12	33.76	18.36
KIDKSIFPGIQGGPLMH	613.73	2	3	1.0526	1.5877	−4.11	12.03	30.43	18.40
QMLRMMMMQMGMKPSQKKINQMMK	1481.93	4	2	1.3223	1.6338	−9.93	49.34	27.48	21.86
RILLSLFLS	531.67	1	2	0.9681	1.5406	−3.02	8.19	29.74	21.55
LTELQVRHII	1222.46	1	1	1.3716	1.4101	−5.64	48.06	23.55	24.51
EPIQSFFQID	1224.34	0	1	1.1369	1.4701	−4.17	51.92	26.60	25.32
NRAVGFISFVI	1223.45	1	1	1.1616	1.5144	−4.62	58.59	27.68	30.91
IHTLEHLLAFTI	1408.67	0	1	1.0082	1.5688	−5.13	81.04	29.04	52.00
GQEQLIPPLIL	1221.47	0	1	1.3977	1.4715	−4.79	81.34	26.17	55.17
PIITVAKEAWPTL	1439.72	1	1	0.9968	1.5364	−6.34	83.17	27.08	56.09
IIGYLDQME	541.63	0	2	1.0583	1.3894	−2.12	83.84	25.56	58.28
IGLLIFLP	886.16	0	1	1.2041	1.5192	1.17	93.97	32.24	61.73
IVLKY	635.82	1	1	0.9641	1.2298	0.91	86.20	22.73	63.47
GIIAAYG	664.77	0	1	1.0865	1.2361	−0.43	89.29	21.92	67.37
PKCPV	543.70	1	1	0.934	0.7843	−2.24	77.72	6.00	71.72
PQTPVP	638.74	0	1	1.1985	0.8503	−2.83	80.37	7.68	72.69
LAAGISTI	745.89	0	1	1.1611	1.2704	−5.24	92.62	19.36	73.26
IDFPTNITMD	1167.31	0	1	1.3316	1.3871	−5.3	96.80	23.07	73.73
DGITDVL	732.80	0	1	1.0216	1.1875	−6.2	93.5	15.96	77.54
HGGSLSAPAIH	1047.15	0	1	1.2372	1.2628	−6.4	97.03	18.23	78.80

times was higher than 1 min, but lower than 3 min. The range from 3 to 5 min in retention time difference was characteristic for 11 peptides, constituting 26.83% of the studied set. The highest numbers of peptides (13) were characterized by 5 to 10 min difference in retention times (31.76%). On the other hand, the highest values, over 10 min, of the difference between predicted and experimental retention times were characteristic for 4 peptides (9.76%) and the largest difference was 11.7 min (Table 4). The correlation between experimental and predicted retention times is still reasonably with correlation coefficient equaled 0.8405 (Figure 3). Some peptides in this set seem to be also false positives (e.g., DQDISGEKATADQLLKDVK or IQNGDPIAGLFDEFTQTVQR), even though they fulfill the established level of X_{corr} criterion for proper peptide identification. The differences between predicted and experimental retention times (here 11.49 and 11.70 minutes, resp.) suggest that these peptides, and proteins, from which they originate, may not be really present in the analyzed sample.
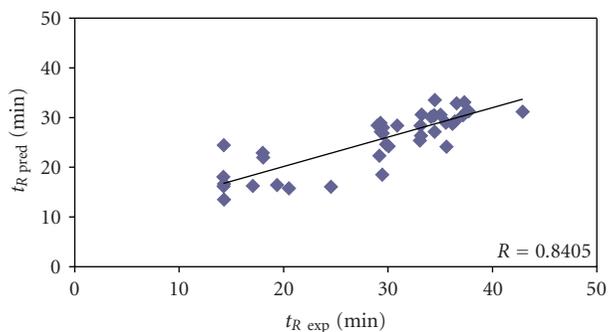
FIGURE 3: Correlation between experimental and predicted retention times for a set of test peptides obtained from *Bacillus subtilis* proteome. The proteins were identified on the basis of one peptide with X_{corr} above 1.5 ($n = 41$).

FIGURE 4: Correlation between experimental and predicted retention times for a set of test peptides obtained from *Bacillus subtilis* proteome. The proteins were identified on the basis of one peptide with X_{corr} below 1.5 ($n = 40$).

Finally, in the group of 40 *Bacillus subtilis* peptides, belonging to proteins identified again on the basis of just one peptide, but with X_{corr} below 1.5 (Table 5), the differences between experimental and predicted retention times range from 1.27 to 78.80 min (mean value equaled 29.41 min). There were only 4 peptides (10%) with predicted and experimental retention times varied less than 3 min. In next 5 cases this difference was over 3 but lower than 5 min, which makes 12.5%. There were 3 peptides (7.5%) in the range between 10 and 15 min of difference in predicted and experimental retention times. For other 5 peptides, the difference in predicted and experimental retention times was from 15 to 20 min (12.5%). Next 4 (10%) peptides in the group belonging to proteins identified on the basis of one peptide with X_{corr} below 1.5 were characterized by 20 to 30 min difference between predicted and experimental retention times. There was 1 case (2.5%), where this difference in retention times ranged between 30 and 50 min. For last 13 peptides (32.5%) in this set the experimental and predicted retention times varied even over 50 min: there were 4 cases (10%), where these values differed between 50 and 60 min; 3 peptides (7.5%) in the 60 to 70 range of retention time difference and 6 (15%) varying more than 70 min (Table 5). It must be stated that for peptides belonging to proteins identified on the basis of one peptide with X_{corr} below 1.5, correlation between experimental and predicted retention times cannot be observed (Figure 4). Therefore it may be concluded that a large number of peptides in this set should be classified as false positives, especially those ones with extremely high difference between experimental and predicted retention times (e.g., HGGSLSAPAIH, DGITDVL, IDFPTNITMD, or LAAGISTI, where these differences are 78.80, 77.54, 73.73, and 73.26 minutes, resp.).

Generally, it can be noticed that lower values of X_{corr} correlate with the higher percentage of peptides are characterized by larger difference between experimental and predicted retention times (Figure 5). In particular, it is observed, when comparing the percentage of cases, where differences between predicted and experimental retention times are higher than 15 min, that in each group of *Bacillus subtilis* peptides belonging to proteins and identified on the
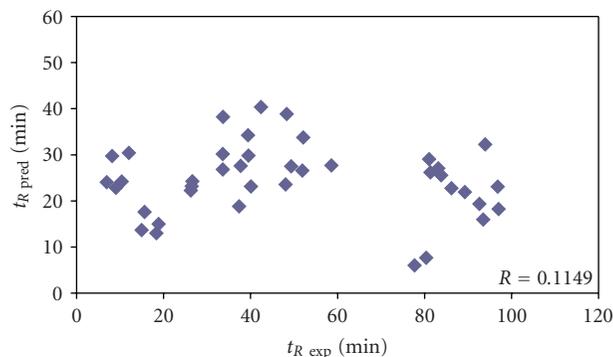
basis of the following: one peptide with X_{corr} below 1.5 (Table 5), one peptide with X_{corr} over 1.5 (Table 4), and more than one peptide with X_{corr} over 1.5 (Table 3). The percentages of peptides characterized by higher than 15 min difference in experimental and predicted retention times in these groups are 57.5%, 0%, and 3.7%, respectively. On the other hand, in model and testing sets of peptides obtained from model proteins all differences between predicted and experimental retention times were lower than 15 min (Tables 1 and 2). It is noticeable that high percent of peptides with low values of X_{corr} was characterized by differences between predicted and experimental retention times larger than 15 min, what can provide an additional indication that they could be considered as potential false positives and in fact were not identified in the analyzed sample. Therefore, QSRR equation to predict peptides retention times might be useful tool to increase throughput of the protein identification in LC-MS/MS.

4. Conclusions

Quantitative structure-retention relationships (QSRRs) model derived with the use of set of peptides identified with the highest scores and originated from 8 known proteins was tested with regards to its predictive capability of the retention time prediction. *Bacillus subtilis* proteome digest was used to check the predictive ability of the novel QSRR model proposed in the study. It was found that the QSRR approach can be applied as the additional constraint in proteomic research verifying results of MS/MS ion search and confirming the correctness of the peptides identifications along with the indication of the potential false positives. The results suggested that due to the QSRR used for the prediction of peptide retention, liquid chromatography separation stage of proteomic research could be useful in the final identification of peptides, especially considering the most uncertain protein identifications based on findings for just one peptide.
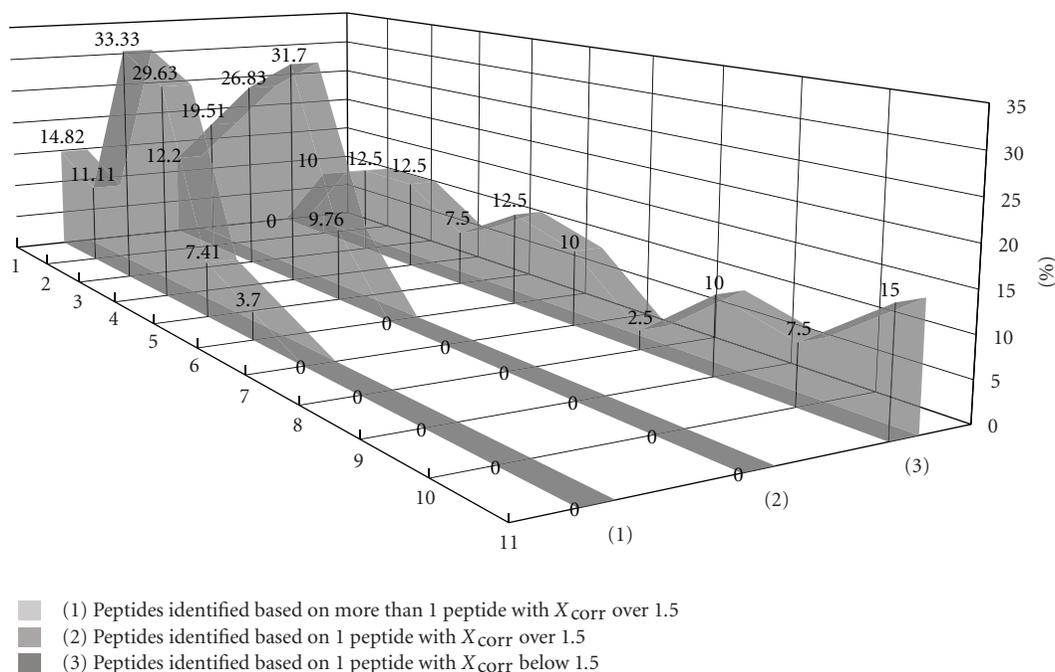
FIGURE 5: Percentage of the difference between predicted and experimental retention times (Dt_R) of *Bacillus subtilis* proteins identified on the basis of one peptide with X_{corr} below 1.5 ($n = 40$), over 1.5 ($n = 41$), and more than one peptide with X_{corr} over 1.5 ($n = 54$).

Acknowledgments

References

[1] T. Fröhlich and G. J. Arnold, "Proteome research based on modern liquid chromatography—tandem mass spectrometry: separation, identification and quantification," *Journal of Neural Transmission*, vol. 113, no. 8, pp. 973–994, 2006.

[2] K. Shinoda, M. Sugimoto, M. Tomita, and Y. Ishihama, "Informatics for peptide retention properties in proteomic LC-MS," *Proteomics*, vol. 8, no. 4, pp. 787–798, 2008.

[3] J. R. Yates III, J. K. Eng, A. L. McCormack, and D. Schieltz, "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database," *Analytical Chemistry*, vol. 67, no. 8, pp. 1426–1436, 1995.

[4] D. C. Anderson, W. Li, D. G. Payan, and W. S. Noble, "A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores," *Journal of Proteome Research*, vol. 2, no. 2, pp. 137–146, 2003.

[5] J. K. Eng, A. L. McCormack, and J. R. Yates III, "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database," *Journal of the American Society for Mass Spectrometry*, vol. 5, no. 11, pp. 976–989, 1994.

[6] D. L. Tabb, W. H. McDonald, and J. R. Yates III, "DTASelect and contrast: tools for assembling and comparing protein identifications from shotgun proteomics," *Journal of Proteome Research*, vol. 1, no. 1, pp. 21–26, 2002.

[7] M. P. Washburn, D. Wolters, and J. R. Yates III, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology," *Nature Biotechnology*, vol. 19, no. 3, pp. 242–247, 2001.

[8] J. Peng, J. E. Elias, C. C. Thoreen, L. J. Licklider, and S. P. Gygi, "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome," *Journal of Proteome Research*, vol. 2, no. 1, pp. 43–50, 2003.

[9] W.-J. Qian, T. Liu, M. E. Monroe, et al., "Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome," *Journal of Proteome Research*, vol. 4, no. 1, pp. 53–62, 2005.

[10] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry," *Nature Methods*, vol. 4, no. 3, pp. 207–214, 2007.

[11] T. Bączek and R. Kaliszan, "Predictions of peptides' retention times in reversed-phase liquid chromatography as a new supportive tool to improve protein identification in proteomics," *Proteomics*, vol. 9, no. 4, pp. 835–847, 2009.

[12] R. C. Dwivedi, V. Spicer, M. Harder, et al., "Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics," *Analytical Chemistry*, vol. 80, no. 18, pp. 7036–7042, 2008.

[13] K. Petritis, L. J. Kangas, B. Yan, et al., "Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information," *Analytical Chemistry*, vol. 78, no. 14, pp. 5026–5039, 2006.

[14] O. V. Krokhin, "Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents," *Analytical Chemistry*, vol. 78, no. 22, pp. 7785–7795, 2006.

[15] R. Kaliszan, "QSRR: quantitative structure-(chromatographic) retention relationships," *Chemical Reviews*, vol. 107, no. 7, pp. 3212–3246, 2007.

[16] R. Kaliszan, *Structure and Retention in Chromatography: A Chemometric Approach*, Harwood, Amsterdam, The Netherlands, 1997.

[17] R. Kaliszan, *Quantitative Structure-Chromatographic Retention Relationships*, John Wiley & Sons, New York, NY, USA, 1987.

[18] J. L. Meek, "Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 3, pp. 1632–1636, 1980.

[19] C. A. Browne, H. P. J. Bennett, and S. Solomon, "The isolation of peptides by high-performance liquid chromatography using predicted elution positions," *Analytical Biochemistry*, vol. 124, no. 1, pp. 201–208, 1982.

[20] V. Casal, P. J. Martin-Alvarez, and T. Herraiz, "Comparative prediction of the retention behaviour of small peptides in several reversed-phase high-performance liquid chromatography columns by using partial least squares and multiple linear regression," *Analytica Chimica Acta*, vol. 326, no. 1–3, pp. 77–84, 1996.

[21] D. Guo, C. T. Mant, A. K. Taneja, J. M. R. Parker, and R. S. Rodges, "Prediction of peptide retention times in reversed-phase high-performance liquid chromatography—I: determination of retention coefficients of amino acid residues of model synthetic peptides," *Journal of Chromatography*, vol. 359, pp. 499–517, 1986.

[22] M. Palmblad, M. Ramström, K. E. Markides, P. Håkansson, and J. Bergquist, "Prediction of chromatographic retention and protein identification in liquid chromatography/mass spectrometry," *Analytical Chemistry*, vol. 74, no. 22, pp. 5826–5830, 2002.

[23] K. Petritis, L. J. Kangas, P. L. Ferguson, et al., "Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses," *Analytical Chemistry*, vol. 75, no. 5, pp. 1039–1048, 2003.

[24] K. Shinoda, M. Sugimoto, N. Yachie, et al., "Prediction of liquid chromatographic retention times of peptides generated by protease digestion of the *Escherichia coli* proteome usina artificial neural networks," *Journal of Proteome Research*, vol. 5, no. 12, pp. 3312–3317, 2006.

[25] E. F. Strittmatter, L. J. Kangas, K. Petritis, et al., "Application of peptide LC retention time information in a discriminant function for peptide identification by tandem mass spectrometry," *Journal of Proteome Research*, vol. 3, no. 4, pp. 760–769, 2004.

[26] R. Kaliszan, T. Bączek, A. Cimochowska, P. Juszczyk, K. Wiśniewska, and Z. Grzonka, "Prediction of high-performance liquid chromatography retention of peptides with the use of quantitative structure-retention relationships," *Proteomics*, vol. 5, no. 2, pp. 409–415, 2005.

[27] T. Bączek, P. Wiczling, M. Marszałł, Y. Vander Heyden, and R. Kaliszan, "Prediction of peptide retention at different HPLC conditions from multiple linear regression models," *Journal of Proteome Research*, vol. 4, no. 2, pp. 555–563, 2005.

[28] T. Bączek, C. Temporini, E. Perani, G. Massolini, and R. Kaliszan, "Identification of peptides in proteomics supported by prediction of peptide retention by means of quantitative structure-retention relationships," *Acta Chromatographica*, no. 18, pp. 72–92, 2007.

[29] M. Michel, T. Bączek, S. Studzińska, et al., "Comparative evaluation of high-performance liquid chromatography stationary phases used for the separation of peptides in terms of quantitative structure-retention relationships," *Journal of Chromatography A*, vol. 1175, no. 1, pp. 49–54, 2007.

[30] I. A. Tarasova, V. Guryča, M. L. Pridatchenko, et al., "Standardization of retention time data for AMT tag proteomics database generation," *Journal of Chromatography B*, vol. 877, no. 4, pp. 433–440, 2009.

[31] O. V. Krokhin, R. Craig, V. Spicer, et al., "An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS," *Molecular and Cellular Proteomics*, vol. 3, no. 9, pp. 908–919, 2004.

[32] M. Gilar, P. Olivova, A. B. Chakraborty, A. Jaworski, S. J. Geromanos, and J. C. Gebler, "Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum," *Electrophoresis*, vol. 30, no. 7, pp. 1157–1167, 2009.

[33] A. Sala, S. Capaldi, M. Campagnoli, et al., "Structure and properties of the C-terminal domain of insulin-like growth factor-binding protein-1 isolated from human amniotic fluid," *Journal of Biological Chemistry*, vol. 280, no. 33, pp. 29812–29819, 2005.

[34] K. Bodzioch, T. Bączek, R. Kaliszan, and Y. Vander Heyden, "The molecular descriptor log SumAA and its alternatives in QSRR models to predict the retention of peptides," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 50, no. 4, pp. 563–569, 2009.

Research Article

Improved Label-Free LC-MS Analysis by Wavelet-Based Noise Rejection

Salvatore Cappadona,[1] Paolo Nanni,[2] Marco Benevento,[2] Fredrik Levander,[3] Piera Versura,[4] Aldo Roda,[2] Sergio Cerutti,[1] and Linda Pattini[1]

[1] *Department of Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy*
[2] *Department of Pharmaceutical Science, University of Bologna, Via Belmeloro 6, 40126 Bologna, Italy*
[3] *Department of Immunotechnology, Lund University, BMC D13, 221 84 Lund, Sweden*
[4] *Ophthalmology Unit, University of Bologna S.Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy*

Correspondence should be addressed to Salvatore Cappadona, salvatore.cappadona@polimi.it

Label-free LC-MS analysis allows determining the differential expression level of proteins in multiple samples, without the use of stable isotopes. This technique is based on the direct comparison of multiple runs, obtained by continuous detection in MS mode. Only differentially expressed peptides are selected for further fragmentation, thus avoiding the bias toward abundant peptides typical of data-dependent tandem MS. The computational framework includes detection, alignment, normalization and matching of peaks across multiple sets, and several software packages are available to address these processing steps. Yet, more care should be taken to improve the quality of the LC-MS maps entering the pipeline, as this parameter severely affects the results of all downstream analyses. In this paper we show how the inclusion of a preprocessing step of background subtraction in a common laboratory pipeline can lead to an enhanced inclusion list of peptides selected for fragmentation and consequently to better protein identification.

1. Introduction

In the last years, clinical proteomics has witnessed an increased interest towards mass spectrometry-based methods for quantitative differential analysis of protein content in biological samples (i.e., biological fluids from drug-treated versus untreated subjects, or from healthy versus ill patients).

MS-based proteomics approaches for comparative analysis include both methods based on the use of stable isotopes [1], such as iTRAQ (Isobaric tags for relative and absolute quantitative) and SILAC (Stable isotope labeling with amino acids in cell culture), and so-called label-free approaches [2].

Label-free liquid chromatography-mass spectrometry (LC-MS) differential analysis allows determining the differential expression level of proteins in multiple samples without presenting any limit to the number of samples being compared and without increasing the complexity of mass spectra. It is based on the direct comparison of peak intensi- ties between multiple runs obtained by continuous detection in MS mode, followed by MS/MS fragmentation of only differentially expressed peptides. This procedure avoids the bias toward abundant peptides, typical of data-dependent tandem MS, and allows an increased identification of low-abundant peptides.

In a typical label-free LC-MS experiment each analysis is performed independently and it is followed by comparison of the multiple LC-MS images. The computational framework includes the steps of peaks detection, maps alignment and normalization, peaks matching across multiple sets, and a statistical analysis of the detected features for the evaluation of the differentially expressed peptides.

Several open-source, commercial, and custom software packages that address one or more of these processing steps have been described in the literature [3]. Nevertheless, most of the available tools show little or no care in assessing a minimum quality standard for the LC-MS maps

entering the pipeline. Baseline subtraction and denoising, for instance, are still often neglected, despite their strong impact on all downstream analyses [4, 5].

In order to show the importance of noise rejection, in this study we report a label-free LC-MS differential analysis of protein abundance in tears samples performed with and without inclusion of a preprocessing step [6] into an established analytical and computational strategy [7, 8].

The preprocessing is performed on a whole LC-MS map. The algorithms work iteratively by first extracting all Single Ion Chromatograms (SICs) and by then processing independently each SIC by means of a wavelet decomposition to identify and remove the components of the chemical and the random noise.

Several other papers have introduced algorithms that exploit the two-dimensional nature of the data to minimize the noise in the mass domain by signal processing in the chromatographic time domain. The advantage of our denoising strategy over other algorithms, though, mainly comes from characterising and subtracting the noise features from all SICs independently. The limit of other common approaches like CODA [9] or MEND [10], in fact, is that they often process only a selection of SICs. CODA, for instance, automatically retains only chromatograms with high S/N ratio and combines them to form a reduced total ion chromatogram (TIC) trace. Similarly, MEND divides the whole mass range in consecutive regions and for each region it determines a global model of the noise by combining a fixed number of "vacant" SICs, that contain no chromatographic peaks.

In the first step of the present work, tears proteins from healthy (H) subjects and patients affected by hyperevaporative dry eye (HDE) were subjected to tryptic digestion and analyzed by reverse-phase chromatography nano-LC ESI-QTOF MS in order to evaluate the protein changes related to the disease. The msInspect software [11] was used for alignment and normalization of the LC-MS maps while the open-source Proteios Software Environment (ProSE) [12] was used for statistical analysis and for the creation of the list of peptides to be identified by RP nano-LC ESI/QTOF MS/MS analysis followed by database search.

In a second phase of the work the same LC-MS raw data files were first filtered to remove chemical and random noise and then reprocessed by the same computational pipeline. The results are compared with those obtained by means of the standard procedure and the influence of noise rejection on the selection of peptides for MS/MS fragmentation is commented according to previously obtained outcomes [13].

2. Experimental

2.1. Materials and Reagents.
2.1. Materials and Reagents. All the analytical grade reagents, the Myoglobin, and the solvents were purchased from Sigma Aldrich (St. Louis, MO, USA).

2.2. Subjects Studied.
2.2. Subjects Studied. A total of 4 subjects, including 2 healthy volunteers (H; 1 M and 1 F) and 2 hyperevaporative dry eye patients (HDE; 2 M), were admitted to this study.

Inclusive criteria for patients were Schirmer test *I* value ≥ 10 mm/5 min, Tear Film Break Up Time (T.F.B.U.T.) < 10 seconds, and symptoms of ocular discomfort from at least two months. Inclusive criteria for healthy control subjects were Schirmer test *I* value ≥ 10 mm/5 min, T.F.B.U.T. ≥ 10 seconds, and no ocular discomfort symptoms. In both groups exclusion criteria were considered the presence of punctuate cheratopaty and/or autoimmune diseases, the use of contact lenses and any ocular surgery in the last 6 months.

All the tear samples were provided by the Ophthalmology Unit at the University of Bologna (Italy) after obtaining informed consent from the subjects studied and according to DEWS guide lines [14] A minimum of $5 \mu L$ tears was collected using a micropipette with sterile tip, centrifuged and stored as previously described [15].

2.3. Tear Samples Preparation.
2.3. Tear Samples Preparation. Total protein quantification of each tear sample was performed by Bradford protein assay using bovine serum albumin (BSA) as a standard according to the manufacturers' instructions (Bio-Rad, Laboratories Inc., CA, USA).

For each sample, $10 \mu g$ of proteins were diluted to $10 \mu L$ with 6 M Urea in 100 mM ammonium bicarbonate pH 8.2 and 2 picomoles of a Myoglobin (Myo, P68082) were added as internal standard. The protein mixtures were reduced by adding $1 \mu L$ of 100 mM dithiothreitol (DTT, Sigma) in 100 mM ammonium bicarbonate for 1 hour at 37°C and alkylated by addition of $3 \mu L$ of 100 mM iodoacetamide (IAA, Sigma) in 100 mM ammonium bicarbonate for 1 hour at room temperature in the dark. The resulting samples were incubated overnight at 37°C with trypsin 12 ng/μL (Promega) in a 50 : 1 (w : w) ratio. The tryptic digestions were blocked after 15 hours incubation with $1 \mu L$ of formic acid, and afterwards the samples were lyophilized to dryness and resolubilized with $20 \mu L$ of 0.1% formic acid (FA).

2.4. Liquid Chromatography and Mass Spectrometry.
2.4. Liquid Chromatography and Mass Spectrometry. For each sample $4 \mu L$ were analyzed by LC-MS analysis using a CapLC (Waters, Manchester, U.K) with flow splitting from $5 \mu L$/min to 200 nL/min, connected with a nanoelectrospray interface to a QTOF Ultima (Waters) using MassLynx v4.0 software as operating software. The peptide separation was performed on an Atlantis dC18 NanoEase column (150×0.075 mm, 3μm) (Waters) with an Atlantis dC18 NanoEase precolumn (0.3×5 mm, 5μm particle size) (Waters), using as mobile phase A H_2O/acetonitrile (95 : 5) 0.1% FA while the mobile phase B was acetonitrile/H_2O (95 : 5) 0.1% FA. A 90-minute chromatographic gradient was used to give a linear increase after 3 minutes from 2% B to 35% B in 70 minutes and from 35% B to 80% B in 2 minutes, and after 3 minutes at 80% B the column is conditioned again at 2% B for 15 minutes. One blank injection with a 30-minute gradient was run between samples to reduce sample carry over, and every six samples 2 pmol of Myo tryptic digest were analyzed as quality control using the same 90 minutes gradient to evaluate the experimental variation. During MS analysis the QTOF was set to scan in profile mode m/z 400–1800 with

(a)

(b)
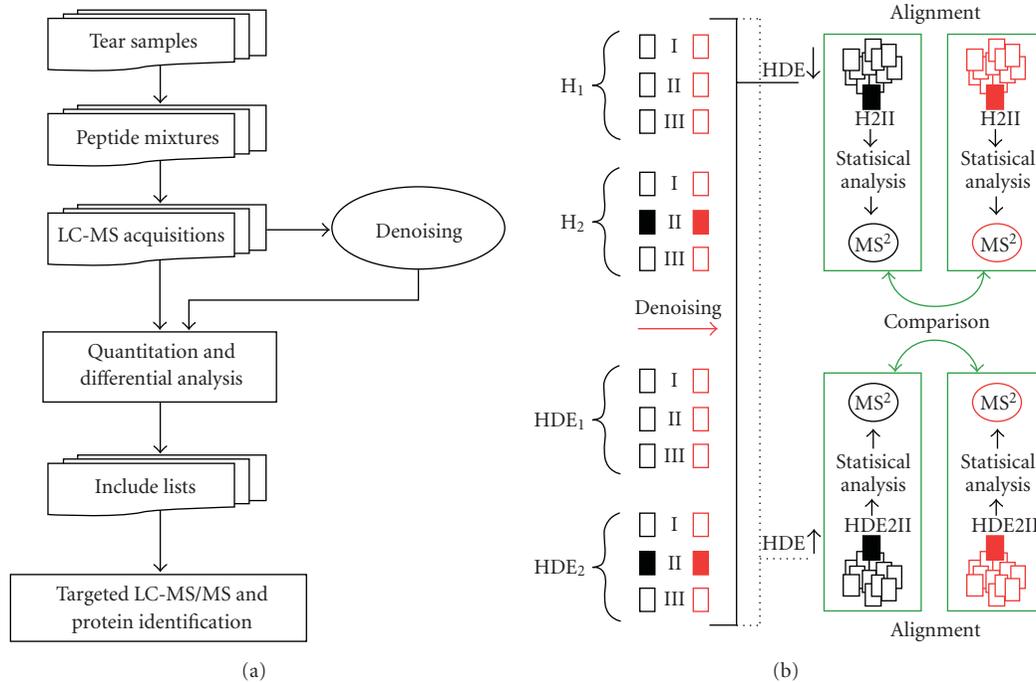
FIGURE 1: Typical pipeline for comparative LC-MS analysis. (a) The preprocessing step of noise rejection is shown as optional to the common workflow. (b) Details of the bioinformatic steps: in black the standard workflow, in red the alternative workflow subsequent to noise rejection, and in green the comparison of the results obtained by the two approaches.

1.9 seconds per scan and 0.1 seconds of scan delay. The samples were analyzed in triplicates. Four microliters of sample were injected for targeted MS/MS and the same LC gradient was used. The survey scan time was set to 1 second and a peak limit of 15 counts to switch to MS/MS mode. For inclusion lists the time tolerance was set to 300 seconds.

2.5. MS Data Analysis. The pipeline of MS data analysis has been already described elsewhere by our research group [7, 8] and is shown in Figure 1(a). Briefly, massWolf (v2.0, http://sashimi.sourceforge.net) was used for the conversion of Micromass LC-MS raw data file to mzXML, while the peptide feature detection was performed using msInspect v2.1. Two alignments were performed for the analysis of all the LC-MS features: one with an H sample and one with an HDE sample as master. The normalization of the LC-MS maps and their alignments were performed by means of the peptide Array tool in msInspect [11], using a mass window of 0.2 m/z and a time window of 250 scans. After alignment, significantly upregulated features were scheduled for targeted MS/MS in inclusion lists generated using the ProSE 2.1 platform [12]. The inclusion limit was a fold change of at least 1.5 and a P-value of 0.05 in a Student's t-test. For the t-test the total intensities, which represent the integrated peak volumes, were used. For features where peaks could not be found in the healthy control samples, a value of 50 ion counts was used, which was an estimate for the detection level in the present setup. Selected features were sorted according to intensity and put into include lists with a maximum of 300 peaks per include list. The retention

times of the second technical replicate acquired sample were used in the include list. Targeted MS/MS analysis was finally performed to identify the peptides contained in the include lists.

2.6. Tandem MS Data Analysis. To generate peak lists for peptide identification, ProteinLynx Global Server 2.2 (Waters) was used. The XML format peak lists were converted to mzData using ProSE. Mascot version 2.2 (www.matrixscience.com) was used for peptide identification.

The Sprot human database, version 57.3, was used, 468851 sequences in total. The search settings were 0.2 Da precursor and 0.6 Da fragment tolerances, carbamidomethylation of cysteine as fixed modification, methionine oxidation as variable modification, and semiTrypsin with one missed cleavage as enzymatic digestion. The search results were exported as XML and matched with MS features using a ProSE plug-in, with a retention time tolerance of 100 seconds and a mass tolerance of 0.12 Da. Proteins were considered correctly identified when at least two different peptides (with significant individual score, i.e., $P < 0.05$) were present.

2.7. Noise Rejection. In the second phase of this work, the original mzXML files entering the pipeline described in Figure 1(a) were cleaned from extraneous noise by a wavelet-based algorithm already described by our research group [6]. Briefly, the algorithm works on a whole LC-MS map by first extracting all Single Ion Chromatograms (SICs) from the spectrographic data and by then decomposing each SIC to

TABLE 1: Alignment of technical replicates. Mass Window = 0.2 Th, Scan Window = 250 scans, Min Charge = 2^+.

		Total	Only I	Only II	Only III	I-II	I-III	II-III	I-II-III
H1 I-II-III	*UNPROCESSED*	4167	767	568	564	315	288	208	1457
	PROCESSED	5736	1061	915	1018	379	361	338	1664
	% increase	**37.65**	**38.33**	**61.09**	**80.50**	**20.32**	**25.35**	**62.50**	**14.21**
H2 I-II-III	*UNPROCESSED*	4476	890	933	625	208	420	347	1053
	PROCESSED	5783	1178	1307	925	228	519	472	1154
	% increase	**29.20**	**32.36**	**40.09**	**48.00**	**9.62**	**23.57**	**36.02**	**9.59**
HDE1 I-II-III	*UNPROCESSED*	4456	861	889	607	314	475	245	1065
	PROCESSED	6324	1301	1386	1051	384	650	347	1205
	% increase	**41.92**	**51.10**	**55.91**	**73.15**	**22.29**	**36.84**	**41.63**	**13.15**
HDE2 I-II-III	*UNPROCESSED*	3921	842	543	563	345	253	288	1087
	PROCESSED	5087	1068	754	962	356	299	394	1254
	% increase	**29.74**	**26.84**	**38.86**	**70.87**	**3.19**	**18.18**	**36.81**	**15.36**
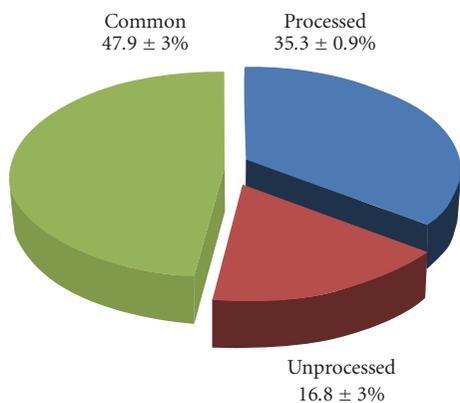
FIGURE 2: Results of the alignment of the original and the filtered files of all 12 LC-MS run. Common: features found in both maps; processed: features found only after noise rejection; unprocessed: features found only in the original maps.

FIGURE 3: KL ratio of corresponding peptides after the alignment of the processed and the unprocessed maps of HDE2III. Most of the peptides have $KL_{Clean} < KL_{Original}$.

identify and remove the noise components. This cleaning step is simply added to the standard pipeline (Figures 1(a) and 1(b)) to selectively remove random and chemical noise while leaving the peptide peaks unaffected.

All data were processed by a stand-alone Java application on a 2.66 GHz iMac running Mac OS X with 1 GB of RAM allocated for the JVM heap. The decomposition was performed on 6 scales and by means of the Coifmann wavelet of degree 1.

3. Results and Discussion

3.1. Effects of Denoising on Peptide Feature Detection. The peptide Array tool of msInspect was run on the technical replicates of all subjects and the results are shown in Table 1. The first column of the table shows that the number of peptide features detected by msInspect increases on average by about 35%. In order to access sensitivity of the cleaning, only peptides with a charge ≥2 were included in the table. Uncharged features were excluded to avoid spurious peaks, while singly charged peptides were excluded to avoid false positive identifications caused by chemical noise, whose regular pattern of peaks occurring at every Th can easily be mistaken for the isotopic distribution of a 1^+ peptide. Furthermore, by only considering peptides that are aligned through at least two of the replicates or through all of the replicates, the average increase in detection can be estimated, respectively, around 20% or 13%. A second alignment was performed between the original and the filtered files of each LC-MS run. In this case, since each file was practically aligned to itself, stricter parameters of Mass Window = 0.05 Da and Scan Window = 5 scans were imposed. Figure 2 shows that on average half of the peptides are found both before and after the cleaning. The figure shows also that about one sixth of the peptides have disappeared from the original file because of the cleaning, while one third have emerged in the processed files after background subtraction.

The biggest region of the pie chart relates to the peaks that are left unaffected by the cleaning, typically high-intensities peptides. A look into the quality of these peptide features shows a 30% average increase in the number of consecutive scans in which a peptide is detected (data not shown). This improvement is usually obtained by unveiling the lowest peaks of its isotopic distribution.

(a)

(b)
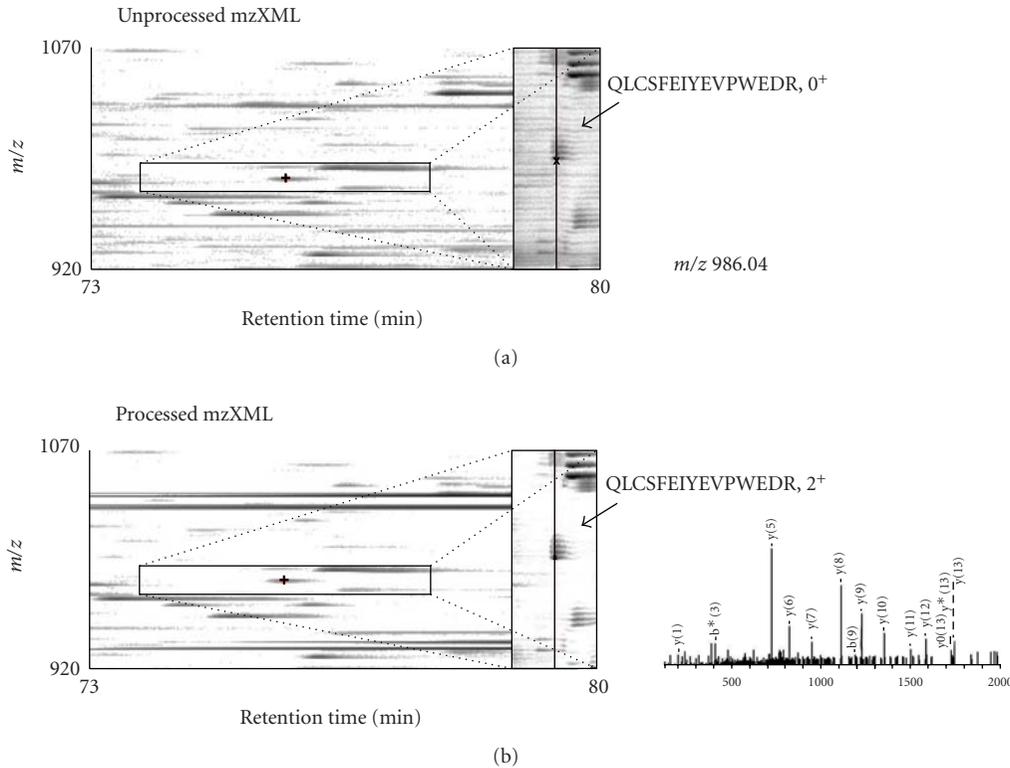
FIGURE 4: Details of unprocessed and processed LC-MS maps. Despite its high intensity, the highlighted doubly charged peptide QLCSFEIYEVPWEDR (m/z 986.04) was not assigned a charge in the original map, because some of its isotopes were altered by the chemical noise.

The general improvement of the quality of the common peptides is also evident in Figure 3. The x-axis gives the average intensity of two aligned peaks, while the y-axis shows the logarithm of their KL ratio. KL is the Kullback-Leibler score and it is used in msInspect as a measure of how closely the observed isotopic distribution of a peptide feature matches its own theoretical distribution at a given mass. This score is always nonnegative and approaches zero for better matches. Therefore all peptides which benefit from a correct denoising will gain a lower KL, will have a ratio $KL_{cl}/KL_{or} < 1$, and will be located below zero on the logarithmic axis. On the opposite, the peptides located above zero are those negatively affected by noise rejection, while points close to the horizontal axis indicate unaffected scores.

The figure thus shows that most of the peptides gained a better (lower) score and that the best improvements were obtained at very low intensities, therefore by those peptides originally masked or hidden by the chemical noise.

The second and the third sectors of the chart show peptides which have appeared or disappeared in the processed maps as a consequence of the preprocessing step. This result is more difficult to interpret because of the high complexity of the samples, which makes it almost impossible to infer which of the found peptides were true positives and which of the lost peptides were true negatives. Nevertheless, visual comparison of original versus clean maps confirms the detection of new low-intensity peaks [6], belonging to sector 2. It also shows that most of the peptides lost in sector 3

are uncharged features which are assigned a charge only after cleaning, like the peptide highlighted in Figure 4.

3.2. Effects of Denoising on Protein Identification. The ProSE toolkit was used to identify over-expressed and under-expressed peptides in the maps of the HDE subjects compared to the H controls. The same statistical analysis was then repeated for the processed data, thus producing a total of 4 include lists (IL), subsequently used for peptide identification by tandem MS (Figure 1(b)).

The results of the MS/MS analysis are summarized in Table 2. Since no processing is performed on tandem MS spectra, the quality of protein identification can be directly ascribed to the quality of the include lists. Despite an expected variability of the identified peptides, which is mostly related to experimental conditions of the targeted acquisitions, the general trend shows a clear improvement of protein identification obtained from the processed data.

Considering the protein under-expressed in HDE patients, the denoising strategy allowed the identification of 4 new proteins and achieved a higher sequence coverage and a better protein score for three proteins already found with the unprocessed data. In particular, Secretoglobin (SG1D1) had 0 peptides in the raw IL and 5 in the clean one, Mammaglobin-B (SG2A1) had 0 versus 3, Ig kappa chain C region (IGKC) had 0 versus 2, Cystatin (CYTS) had 0 versus 4, while Lipocalin-1 (LCN1) had 6 versus 12, Proline-rich

TABLE 2: List of the peptides identified using the include lists from unprocessed and processed data. Bold italic proteins indicate abundance variations in agreement with previous outcomes [13].

(a) Over-expressed proteins in hyperevaporative dry eye patients

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
		590.89	2	GVANALAHKYH	16	18
		862.91	3	GTFATLSELHCDKLHVDPENFR	40	59
		637.89	2	LLVVYPWTQR	31	30
		737.508	3	TLSELHCDKLHVDPENFR	51	—
		725.44	2	VVAGVANALAHKYH	33	27
		835.59	2	VLGAFSDGLAHLDNLK	—	74
Hemoglobin subunit bete	HBB_HUMAN	1030.15	2	FFESFGDLSTPDAVMGNPK	—	17
		888.95	2	LLGNVLVCVLAHHFGK	—	49
		899.64	2	KVLGAFSDGLAHLDNLK	—	53
		575.39	2	VVAGVANALAHK	—	28
		524.805	2	VVYPWTQR	—	14
		810.203	3	FATLSELHCDKLHVDPENFR	—	15
		626.39	2	AGVANALAHKYH	—	17
		786.03	2	LGAFSDGLAHLDNLK	—	22
				Total number of peptides	**5**	**13**
				Protein score	**86**	**169**
		756.49	2	VPQVSTPTLVEVSR	47	—
		613.88	2	FKDLGEENFK	5	17
		717.46	2	KECCEKPLLEK	51	23
Serum albumin	ALBU_HUMAN	820.49	2	KVPQVSTPTLVEVSR	—	85
		722.35	2	YICENQDSISSK	—	30
		871.95	2	HPYFYAPELLFFAK	—	35
		926.8	3	LVRPEVDVMCTAFHDNEETFLKK	—	29
				Total number of peptides	**3**	**6**
				Protein score	**71**	**96**
		523.96	3	FPHFDLSHGSAQVK	—	2
		663.37	2	HFDLSHGSAQVK	—	30
Hemoglobin subunit alpha	HBA_HUMAN	942.98	3	DALTNAVAHVDDMPNALSALSDLHAHK	—	46
		785.92	4	KVADALTNAVAHVDDMPNALSALSDLHAHK	—	6
				Total number of peptides	**—**	**4**
				Protein score	**—**	**55**

(b) Under-expressed proteins in hyperevaporative dry eye patients

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
Lipocalin-1	LCN1_HUMAN	669.86	2	TDEPGKYTADGGK	34	48
		722.94	2	YCEGELHGKPVR	16	—
		707.44	3	DHYIFYCEGELHGKPVR	97	36
		643.64	4	SHVKDHYIFYCEGELHGKPVR	30	15
		874.54	2	NNLEALEDFEKAAGAR	67	84
		934.6	2	YIFYCEGELHGKPVR	14	—
		496.84	2	GELHGKPVR	—	10
		561.36	2	EGELHGKPVR	—	20
		593.88	2	AVLEKTDEPGK	—	28
		641.38	2	CEGELHGKPVR	—	2
		796.48	2	FYCEGELHGKPVR	—	36
		885.5	2	SDEEIQDVSGTWYLK	—	56
		1171.21	2	HHLLASDEEIQDVSGTWYLK	—	102
		857.88	3	SHVKDHYIFYCEGELHGKPVR	—	47
				Total number of peptides	**6**	*12*
				Protein score	**136**	*257*
Polymeric immunoglobulin receptor	PIGR_HUMAN	703.95	2	VLDSGFREIENK	38	44
		876.04	2	TVTINCPFKTENAQK	11	—
		854.01	2	QSSGENCDVVVNTLGK	78	47
		573.35	2	QGARGGCITLI	—	3
		775.97	2	TINCPFKTENAQK	—	6
				Total number of peptides	**3**	**4**
				Protein score	**81**	**63**
Lactotransferrin	TRFL_HUMAN	694.94	2	GPQYVAGITNLKK	16	—
		624.37	2	WCAVGEQELR	15	—
		790.48	2	NLLFNDNTECLAR	37	—
		874.03	2	RSVQWCAVSQPEATK	74	—
		630.91	2	GPQYVAGITNLK	—	21
		702.89	2	QWCAVSQPEATK	—	9
		706.93	2	PIQCIQAIAENR	—	62
		841.97	2	CSTSPLLEACEFLR	—	70
		886.06	2	RDSPIQCIQAIAENR	—	6
		953.55	3	SQQSSDPDPNCVDRPVEGYLAVAVVR	—	30
		982.55	2	SASCVPGADKGQFPNLCR	—	56
				Total number of peptides	**4**	*8*
				Protein score	**77**	*141*
Serum albumin	ALBU_HUMAN	1020.23	2	VFDEFKPLVEEPQNLIK	38	—
		773.99	2	LKECCEKPLLEK	35	—
				Total number of peptides	**2**	—
				Protein score	**46**	—
Ig kappa chain C region	IGKC_HUMAN	714.45	3	HKVYACEVTHQGLSSPVTK	—	79
		1071.18	2	HKVYACEVTHQGLSSPVTK	—	34
				Total number of peptides	—	**2**
				Protein score	—	**79**

(b) Continued.

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
		815.2	2	DRPARHPQEQPLW	43	15
		801.48	2	FPSVSLQEASSFFR	—	84
Proline-rich protein 4	PROL4_HUMAN	537.66	3	PSVSLQEASSFFRR	—	13
		598.38	2	HPPPPPFQNQQRPPR	—	15
		552.68	3	PPPPPFQNQQRPPR	—	27
				Total number of peptides	**1**	**5**
				Protein score	**43**	**84**
		869.57	1	APLEAVAAK	—	24
		572.9	2	FKAPLEAVAAK	—	39
Secretoglobin family 1D member 1	SG1D1_HUMAN	988.67	2	QALGSEITGFLLAGKPVFK	—	96
		712.8	3	CQALGSEITGFLLAGKPVFK	—	71
		622.86	2	KCVDTMAYEK	—	10
				Total number of peptides	*—*	*5*
				Protein score	*—*	*188*
		732.95	2	FKQCFLNQSHR	—	45
		969	2	ELLQEFIDSDAAAEAMGK	—	81
Mammaglobin-B	SG2A1_HUMAN	977.54	2	ELLQEFIDSDAAAEAM̲GK + oxM	—	7
				Total number of peptides	*—*	*3*
				Protein score	*—*	*89*
		964.05	2	PNLDTCAFHEQPELQK	—	21
		685.74	3	PNLDTCAFHEQPELQKK	—	29
Cystatin-S	CYTS_HUMAN	1135.65	2	SQPNLDTCAFHEQPELQKK	—	6
		986.04	2	QLCSFEIYEVPWEDR	—	67
				Total number of peptides	—	**4**
				Protein score	—	**70**

protein 4 (PROL4) had 1 versus 5, and Lactotransferrin (TRFL) had 4 versus 8. One protein was found with a better sequence coverage but a lower protein score (Polymeric Immunoglobulin Receptor (PIGR): 3 versus 4) and a last one was found only before filtering (Serum Albumin (ALBU): 2 versus 0).

Among the proteins over-expressed in HDE patients, 1 protein was found only after noise rejection (Hemoglobin subunit alpha (HBA): 0 versus 4) and 2 proteins achieved a higher sequence coverage and a better score (Hemoglobin subunit beta (HBB): 5 versus 13 and ALBU: 3 versus 6). The obtained results have been compared with the outcomes of a previous study performed by our research group [13], in which the differential expression of proteins in HDE patients over H controls was monitored by mono-dimensional gel electrophoresis and western blot analysis. Under-expressions of LCN1, SG1D1, SG2A1, TRFL, and over-expression of ALBU are in perfect agreement with the previous study and these proteins are shown in bold italic in Table 2. In particular, noise filtering allowed to identify ALBU as over-expressed in HDE patients, whereas the standard pipeline wrongly identified the protein as both under- and over-expressed.

As regards the variations in the abundances of HBA, HBB, IGKC, PROL4, PIGR, and CYTS associated to HDE, this could not be validated by comparison with previously published data. Nevertheless, a clear proof of their proper identification can be observed in Figure 4, where the correct charge assignment of the CYTS peptide QLCSFEIYEVP-WEDR allowed its inclusion in the list of peptides under-expressed in tear samples from HDE patients.

4. Conclusions

We have previously shown that wavelet denoising in the RT domain achieves selective rejection of chemical and random noise while preserving peptides features and morphology. The approach has proved to unveil low-intensity peptides originally masked by the chemical noise and to reduce false positive identification, by filtering noise peaks originally mimicking the peptide morphology.

In this work we have applied our noise filtering strategy to a label-free LC-MS differential analysis of protein abundance in tears samples. The mzXML files have been simply intercepted, processed by our algorithm and reinserted in the standard workflow just before the analyses by msInspect.

The results show that noise rejection allows to increase the sensitivity of msInspect to real peptides and to obtain more accurate include lists for further targeted MS/MS analysis. These results are validated by comparison to previous outcomes which confirm an improvement in terms of number of identified proteins, higher sequence coverage, and better protein scores.

Acknowledgments

References

[1] A. Iliuk, J. Galan, and W. A. Tao, "Playing tag with quantitative proteomics," *Analytical and Bioanalytical Chemistry*, vol. 393, no. 2, pp. 503–513, 2009.

[2] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative mass spectrometry in proteomics: a critical review," *Analytical and Bioanalytical Chemistry*, vol. 389, no. 4, pp. 1017–1031, 2007.

[3] A. H. P. America and J. H. G. Cordewener, "Comparative LC-MS: a landscape of peaks and valleys," *Proteomics*, vol. 8, no. 4, pp. 731–749, 2008.

[4] M. Hilario, A. Kalousis, C. Pellegrini, and M. Müller, "Processing and classification of protein mass spectra," *Mass Spectrometry Reviews*, vol. 25, no. 3, pp. 409–449, 2006.

[5] J. Listgarten and A. Emili, "Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry," *Molecular and Cellular Proteomics*, vol. 4, no. 4, pp. 419–434, 2005.

[6] S. Cappadona, F. Levander, M. Jansson, P. James, S. Cerutti, and L. Pattini, "Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry," *Analytical Chemistry*, vol. 80, no. 13, pp. 4960–4968, 2008.

[7] P. Nanni, L. Mezzanotte, G. Roda, et al., "Differential proteomic analysis of HT29 Cl.16E and intestinal epithelial cells by LC ESI/QTOF mass spectrometry," *Journal of Proteomics*, vol. 72, no. 5, pp. 865–873, 2009.

[8] P. Nanni, F. Levander, G. Roda, A. Caponi, P. James, and A. Roda, "A label-free nano-liquid chromatography-mass spectrometry approach for quantitative serum peptidomics in Crohn's disease patients," *Journal of Chromatography B*, vol. 877, no. 27, pp. 3127–3136, 2009.

[9] W. Windig, J. M. Phalp, and A. W. Payne, "A noise and background reduction method for component detection in liquid chromatography/mass spectrometry," *Analytical Chemistry*, vol. 68, no. 20, pp. 3602–3606, 1996.

[10] V. P. Andreev, T. Rejtar, H.-S. Chen, E. V. Moskovets, A. R. Ivanov, and B. L. Karger, "A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain," *Analytical Chemistry*, vol. 75, no. 22, pp. 6314–6326, 2003.

[11] M. Bellew, M. Coram, M. Fitzgibbon, et al., "A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS," *Bioinformatics*, vol. 22, no. 15, pp. 1902–1909, 2006.

[12] J. Häkkinen, G. Vincic, O. Månsson, K. Wårell, and F. Levander, "The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data," *Journal of Proteome Research*, vol. 8, no. 6, pp. 3037–3043, 2009.

[13] P. Versura, P. Nanni, A. Bavelloni, et al., "Tear protein changes in mild evaporative dry eye," submitted.

[14] "Report of the Dry Eye Workshop (DEWS)," *The Ocular Surface*, vol. 5, no. 2, 2007.

[15] P. Versura, M. Frigato, R. Mulé, N. Malavolta, and E. C. Campos, "A proposal of new ocular items in Sjögren's syndrome classification criteria," *Clinical and Experimental Rheumatology*, vol. 24, no. 5, pp. 567–572, 2006.

Research Article

Proteomic Studies of Cholangiocarcinoma and Hepatocellular Carcinoma Cell Secretomes

Chantragan Srisomsap,[1] Phannee Sawangareetrakul,[1] Pantipa Subhasitanont,[1] Daranee Chokchaichamnankit,[1] Khajeelak Chiablaem,[1] Vaharabhongsa Bhudhisawasdi,[2,3] Sopit Wongkham,[3,4] and Jisnuson Svasti[1,5]

[1] *Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok 10210, Thailand*

[2] *Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand*

[3] *Liver Fluke and Cholangiocarcinoma Research Center, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand*

[4] *Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand*

[5] *Department of Biochemistry, Faculty of Science, Mahidol University, Rama VI Road, Bangkok 10400, Thailand*

Correspondence should be addressed to Chantragan Srisomsap, chantragan@cri.or.th

Received 1 July 2009; Accepted 28 September 2009

Academic Editor: Helen J. Cooper

Cholangiocarcinoma (CCA) and hepatocellular carcinoma (HCC) occur with relatively high incidence in Thailand. The secretome, proteins secreted from cancer cells, are potentially useful as biomarkers of the diseases. Proteomic analysis was performed on the secreted proteins of cholangiocarcinoma (HuCCA-1) and hepatocellular carcinoma (HCC-S102, HepG2, SK-Hep-1, and Alexander) cell lines. The secretomes of the five cancer cell lines were analyzed by SDS-PAGE combined with LC/MS/MS. Sixty-eight proteins were found to be expressed only in HuCCA-1. Examples include neutrophil gelatinase-associated lipocalin (lipocalin 2), laminin 5 beta 3, cathepsin D precursor, desmoplakin, annexin IV variant, and annexin A5. Immunoblotting was used to confirm the presence of lipocalin 2 in conditioned media and cell lysate of 5 cell lines. The results showed that lipocalin 2 was a secreted protein which is expressed only in the conditioned media of the cholangiocarcinoma cell line. Study of lipocalin 2 expression in different types of cancer and normal tissues from cholangiocarcinoma patients showed that lipocalin 2 was expressed only in the cancer tissues. We suggest that lipocalin 2 may be a potential biomarker for cholangiocarcinoma.

1. Introduction

In Thailand, cholangiocarcinoma (CCA), a malignant tumor derived from bile duct epithelium, occurs with a high incidence in tropical countries where it is associated with liver fluke (*Opisthorchis viverrini*) infestation and nitrosamine ingestion [1]. CCA shows high mortality and presents challenges in diagnosis; so prognosis for CCA patients is rather poor. There is still need for better tumor markers for early diagnosis. The most widely used circulating marker for CCA is carbohydrate antigen (CA) 19-9 [2]. However, (CA) 19-9 is also elevated in pancreatic cancer, gastric cancer, and primary biliary cirrhosis and has been shown it gives false positive results [3]. Carcinoembryonic antigen (CEA)

is the other common tumor marker used for detecting CCA. CEA is not specific, being mainly used for colorectal cancers, and can be elevated in other types of cancer, such as gastrointestinal or gynecologic malignancies [4].

Differences in expression profiles between normal liver and CCA tissues were studied, because CCA is contained in liver tissue and is suggested to arise from the same stem cells as HCC [5]. We have previously compared CCA and HCC cell lines using proteomic techniques in order to investigate potential CCA markers for early diagnosis [6]. Comparison of 2D-PAGE patterns for a cholangiocarcinoma cell line (HuCCA-1) and two hepatocellular carcinoma cell lines (HepG2 and HCC-S102) showed that cytokeratin 7 (CK7), cytokeratin 19 (CK19), an unknown proteins

(U2/2), and galectin-3 were found in CCA but not in HCC. An extension of this study investigated membrane proteins and cytosolic proteins [7], which showed that ten membrane proteins were found in HuCCA-1 but not in HCC-S102, including mitogen-activated protein kinase kinase kinase 2, calgizzarin, integrin alpha-6 precursor, ezrin, and hippocalcin-like protein 1. The subproteomic approach used here may be useful for developing potential biomarkers for early detection of CCA. However, proteomic studies using cell lines have limitations, since it is not known whether the differentially expressed proteins will be present in accessible biological fluids such as plasma, serum, or urine or not.

Cells and tissues secrete proteins into the extracellular environment, the secretome, which may reflect a large variety of pathological conditions and may be a useful source of biomarkers. The secreted proteins are known to regulate many biological processes. These proteins are not only components of the extracellular matrix and biological fluids but are also involved in blood coagulation, immune defense, signal transduction, and carcinogenesis [8]. Certain proteins secreted from cancer cells that enter the circulatory system can be utilized as targets for monitoring or screening for the presence of cancer cells. The cancer secretome, such as that from hepatocellular carcinoma [9], lung cancer [10], breast cancer [11], and oral squamous cell carcinoma [12], has been studied by many research groups to determine the release of the total proteins by cancer cells. This technique provides useful tools for the discovery of novel biomarkers, by using a cell culture model system in which the cells were grown in serum-free media for proteomic analysis [13, 14]. Secretome analysis of nasopharyngeal carcinoma (NPC) cell lines has been studied by SDS-PAGE and MALDI-TOF MS and revealed several potential NPC protein markers [15, 16].

Since the secretomes of cholangiocarcinoma have not previously been reported, we have used conditioned media to compare the secretomes of cholangiocarcinoma cell line (HuCCA-1) with those of hepatocellular carcinoma cell lines (HCC-S102, HepG2, SK-Hep-1, and Alexander). Expression of proteins was studied by SDS-PAGE and LC/MS/MS, and overexpression of proteins in HuCCA-1 was confirmed by using 1-DE and 2-DE immunodetection. Finally, the potential biomarker was validated in various types of cholangiocarcinoma tissues.

2. Materials and Methods

2.1. Sample Collection. Tissue samples were collected from Department of Surgery, and Liver Fluke and Cholangiocarcinoma Research Center, Faculty of Medicine, Khon Kaen University. Ethical clearance for the tissues was approved by the Ethics Committee for Human Research of Khon Kaen University (HE471214). After resection, specimens were immediately taken to a pathologist, who sampled both the tumor itself and adjacent normal-appearing bile duct. Both samples were then snap-frozen in liquid nitrogen within 10 minutes of removal from the patients and stored at $-80°C$

until analysis. The details of types, ages, genders, grades, and histopathology are shown in Table 1.

2.2. Cell Cultures. Cholangiocarcinoma cell line, HuCCA-1 derived from bile duct tumor mass of Thai patient, was kindly provided by Professor Sirisinha [17] and grown in Ham's F12 culture medium (Hyclone laboratories, Logan, UT, USA), containing 15 mM HEPES and supplemented with 10% fetal bovine serum (FBS, Hyclone Laboratories), 100 U/mL penicillin, and 100 μg/mL streptomycin and 125 ng/ml amphotericin B. Hepatocellular carcinoma cells, HCC-S102, established from Thai patient [18] was kindly provided by Dr. Sumalee Tungpradakul and grown in RPMI 1640 (Gibco, Grand Island, NY, USA) containing 25 mM HEPES, supplemented with penicillin (100 U/mL), streptomycin (100 μg/mL) amphotericin B (125 ng/mL) (Gibco), and 10% FBS. Another human hepatocellular carcinoma cell line, the Alexander cell line, a human HCC-derived cell line secreting HBsAg, originally obtained from American Type Culture Collection (Rockville, MD, USA) was kindly provided by Dr. Mammen Mammen and Dr. Ananda Nisalak of the Department of Virology, the Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand. The hepatoblastoma cell line, HepG2 and hepatocellular cell line, SK-Hep-1 were purchased from American Type Culture Collection. The Alexander HepG2, and SK-Hep-1 cell lines were grown in DMEM (Gibco) with 10% FBS, 100 U/mL penicillin, 100 μg/mL Streptomycin, and 125 ng/mL amphotericin B. All cells were maintained at 37°C in a humidified atmosphere, 95% air, 5% CO_2 at 37°C.

2.3. Cell Viability Assay. After incubating the cells in complete media or serum-free media for 24 hours, the numbers of viable cells and dead cells were counted by using the trypan blue dye exclusion method. The percentage of cell viability was expressed as the ratio of total viable cells to the sum of total viable and dead cells.

2.4. Harvesting Conditioned Media. To obtain culture supernatants, approximately 3×10^7 cancer cells were grown to 80% confluence and then the cells were washed with serum-free medium 2 times before incubation in serum-free medium for 24 hours. After incubation, the conditioned medium was harvested and centrifuged 800x*g* at 4°C for 10 minutes to remove suspended cells. The supernatant was dialyzed against distilled water using a dialysis membrane with molecular weight cutoff 3500 Da (Cellu Sep, Texas, USA) for 48 hours and then concentrated by SpeedVac. The concentrations of total proteins were determined by the Bradford protein assay reagent (Bio-Rad Laboratories, CA) [19].

2.5. SDS-PAGE of Conditioned Media from Cell Lines. Twenty micrograms of conditioned media from five cell lines were resolved on 8%–14% gradient SDS-PAGE ($100 \times 105 \times 0.75$ mm). Electrophoresis was performed in a Hoefer system (Hoefer, Inc., San Francisco, CA, USA) at 12 mA, room temperature for 2 hours. The protein was stained with

Coomassie brilliant blue R-250 (SERVA Electrophoresis GmbH, Heidelberg, Germany).

2.6. Western Blotting and Immunodetection.

Protein extracts, separated by 12% SDS-PAGE and transferred onto nitrocellulose membrane (Hybond ECL, GE Healthcare, Buckinghamshire, UK), were probed with monoclonal antibodies against NGAL (Santa Cruz Biotechnology, CA, USA). Proteins of interest were detected using the enhanced chemiluminescence (ECL) plus detection system, with high-performance film (Hyperfilm ECL; GE Healthcare).

2.7. Two-Dimensional Gel Electrophoresis (2-DE).

2D PAGE was performed using the immobiline/polyacrylamide system. Three hundred micrograms of conditioned media from HuCCA-1 were applied by overnight in-gel rehydration of 70 mm, using nonlinear pH 3–10 IPG gel strips (GE Healthcare). The first dimension (IEF) was performed to give a total of 8000 Vh on an IPGphor (GE Healthcare). Focused IPG strips were equilibrated for 10 minutes in a solution (6 M urea, 30% glycerol, 1% SDS 50 mM Tris-HCl buffer (pH 6.8), and 1% DTT), and then for an additional 10 minutes in the same solution except that DTT was replaced by 2.5% iodoacetamide. The IPG strips were then applied to the second dimension 12% T SDS-PAGE ($100 \times 105 \times 1.5$ mm). After electrophoresis, proteins were visualized by Coomassie blue R-250 staining.

2.8. Sample Preparation for SDS-PAGE Immunodetection.

Tissues were homogenized in phosphate buffered saline containing 0.2 mM phenylmethanesulfonylfluoride, 2 μg/mL pepstatin A, 1 μg/mL bestatin, and centrifuged for 20 minutes at 15 000 rpm. Protein content of the supernatant was determined using the Bradford method [19]. Samples (10 μg protein) were mixed with sample buffer, boiled, and applied to 15% T SDS polyacrylamide gels ($100 \times 80 \times 0.75$ mm). Electrophoresis was performed in a Hoefer system (Hoefer, Inc., San Francisco, CA, USA) at 10 mA, room temperature for 1.5 hours, followed by electroblotting of proteins from the gel onto PVDF membranes (Immobilon-P; Millipore, Billerica, MA, USA) at 100 V for 30 minutes at 4°C. After blocking in 5% nonfat dry milk, membranes were probed with 1 : 200 diluted anti-NGAL monoclonal antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA), repeatedly washed in 20 mM Tris buffered saline, pH 7.6, containing 0.1% Tween 20, and then incubated in 1 : 5000 rabbit antimouse immunoglobulin G (IgG; Dako Cytomation, Glostrup, Denmark) for 1 hour. After washing, the reaction was developed using the ECL plus detection system, with high-performance film (Hyper-film ECL; GE Healthcare).

2.9. Tryptic In-Gel Digestion.

Protein spots were excised and transferred to 0.5 mL microfuge tubes. Fifty μL of 0.1 M NH$_4$HCO$_3$ in 50% acetonitrile was added. The gel was incubated 3 times for 20 minutes at 30°C. The solvent were discarded and gel particles were dried completely by Speed Vac. Reduction and alkylation were performed by swelling the gel pieces in 50 μL buffer solution (0.1 M NH$_4$HCO$_3$, 10 mM DTT and 1 mM EDTA) and incubating at 60°C for 45 minutes. After cooling, the excess liquid was removed and quickly replaced by the same volume of freshly prepared 100 mM iodoacetamide in 0.1 M NH$_4$HCO$_3$ solution. The reaction was incubated at room temperature in the dark for 30 minutes. The iodoacetamide solution was removed and the gel pieces were washed with 50% acetonitrile in water, 3 times for 10 minutes each time, and the gel pieces were completely dried. Aliquots of trypsin (Promega Corporation, WI, USA) (1 μg trypsin /10 μL 1% acetic acid) were prepared and stored at -20°C. Fifty μL of digestion buffer (0.05 M Tris HCl, 10% acetonitrile, 1 mM CaCl$_2$, pH 8.5) and 1 μL of trypsin were added to the gel pieces. After incubating the reaction mixture at 37°C overnight, the digestion buffer was removed and saved. The gel pieces were then extracted by adding 60 μL of 2% freshly prepared trifluoroacetic acid and incubating for 30 minutes at 60°C. The extract and the saved digestion buffer were finally pooled and dried.

2.10. Protein Identification by LC/MS/MS.

LC/MS/MS analyses were carried out using a capillary LC system (Waters) coupled to a Q-TOF mass spectrometer (Micromass, Manchester, UK) equipped with a Zspray ion source working in the nanoelectrospray mode. Glu-fibrinopeptide was used to calibrate the instrument in MS/MS mode. The tryptic peptides were concentrated and desalted on a 75 μm ID $\times$ 150 mm C18 PepMap column (LC Packings, Amsterdam, The Netherlands). Eluents A and B were 0.1% formic acid in 97% water, 3% acetonitrile, and 0.1% formic acid in 97% acetonitrile, respectively. Six μL of sample was injected into the nanoLC system, and separation was performed using the following gradient: 0 minute 7% B, 35 minutes 50% B, 45 minutes 80% B, 49 minutes 80% B, 50 minutes 7% B, and 60 minutes 7% B. The database search was performed with ProteinLynx screening SWISS-PROT and NCBI. For some proteins that were difficult to find, the Mascot search tool available on the Matrix Science site screening NCBInr was used. The search parameters were set as follows: peptide mass tolerance = 1 Da; MS/MS ion mass tolerance = 1 Da; enzyme set as trypsin and allowance was set up to two missed cleavages; peptide charges were limited to 2+ and 3+. The proteins were identified with P-values $\leq$ .05 and Mascot scores >35 were considered as promising hits .

3. Results

3.1. Comparison of Secretomes from Cholangiocarcinoma and Hepatocellular Carcinoma Cells.

The secreted proteins from conditioned media of HuCCA-1, HCC-S102, HepG2, SK-Hep-1, and Alexander cells were resolved by SDS-PAGE and visualized by Coomassie blue as shown in Figure 1. The secreted protein patterns were enriched in the conditioned media and showed patterns that differed significantly in each cell line. The effect of 24-hour serum starvation on the viability of the five cancer cells cultured was tested and the results showed that this treatment had little effect on cell viability (Figure 2).

TABLE 1: Medical diagnosis of cholangiocarcinoma patients with tumor types and histopathology.

Sample ID	Gender	Age	Tumor type	Staging	Histopathology
1	F	62	Peripheral, Single, Seg 7	4B	Well differentiated adenocarcinoma, CCA, invade intrahepatic vein
2	F	67	Combined, Single, Seg 7, Hilar, long	4A	Well differentiated tubular type, CCA
3	M	64	Peripheral, Single, Seg 8	4B	Well differentiated tubular type, CCA
4	M	43	Peripheral, Single, Seg 6	4B	Moderately diff adenocarcinoma invade intrahepatic vessel
5	M	65	Peripheral, Multiple, Seg 1-2	4A	Moderately diff, CCA
6	M	46	Peripheral, Single, Seg 6	4A	Moderately diff adenocarcinoma invade intrahepatic vein
7	M	39	Peripheral, Single, Seg 6	4B	Moderately diff, CCA
8	F	56	Peripheral, Single, Seg 6	4B	Poorly diff CCA, invade intrahepatic vein
9	M	67	Combined, Single, Seg 8, Hilar	4A	Poorly diff adenocarcinoma
10	M	53	Central, Hilar, long	3	Papillary mucinous carcinoma
11	F	35	Peripheral, Single	4B	Papillary mucinous cyst adenocarcinoma
12	M	49	Peripheral, Multiple, Seg 2-4	4A	Papillary mucinous carcinoma

Because of the high sensitivity of mass spectrometry in identifying mixtures of proteins, the stained gels were marked and cut into 0.5 mm slices from the gel top to dye front as shown in Figure 1. Forty-eight total gel bands were subjected to in-gel tryptic digestion and the proteins were identified by LC/MS/MS. There are 83 proteins secreted from the HuCCA-1 cell line as shown in Table 2. The secreted proteins from 4 hepatocellular carcinoma cell lines (HCC-S102, HepG2, SK-Hep-1, and Alexander) were collected in the form of a database, together with those of other cell lines including lung, breast, cervical, and oral cell lines, used in our group (manuscript in preparation). The secreted proteins from 4 hepatocellular carcinoma cell lines are shown in Table 1 in supplementary Material available online at doi:10.1155/2010/437143. The molecular weights shown in Table 2 and Supplementary Table 1 are theoretical values from databases. Some experimental molecular weights differed from theoretical values because the bands were from SDS-PAGE, in which the proteins were denatured and found as isozymes. From Table 2 and Supplementary Table 1, some

proteins containing one matched peptide were considered because the peptide score was greater than 35, higher than the Mascot score of 20, which Kristiansen et al. considered to be significant [20]. We also manually selected only the peptide match that had a length of at least 8 amino acids with sequence tag of at least three amino acids to be a good y-ion series. Relevant information on the functions of proteins was taken from Swiss-Prot and NCBI. Then proteins were classified by their functions into various categories, namely, chaperone/stress response, cell cycle, cytoskeleton/mobility, DNA replication/gene regulation, extracellular matrix, immunological response, ion channels, metabolism, protection and detoxification, protein synthesis and degradation, signal transduction, transport/binding proteins, tumor-associated proteins, and unannotated/function inferred.

When the HuCCA-1 cell line is compared to the 4 hepatocellular carcinoma cell lines, all cell lines secreted distinct protein profiles. We also found distinct secreted protein profiles when we compared the secreted proteins to different types of cell lines (liver, lung, breast, cervical and
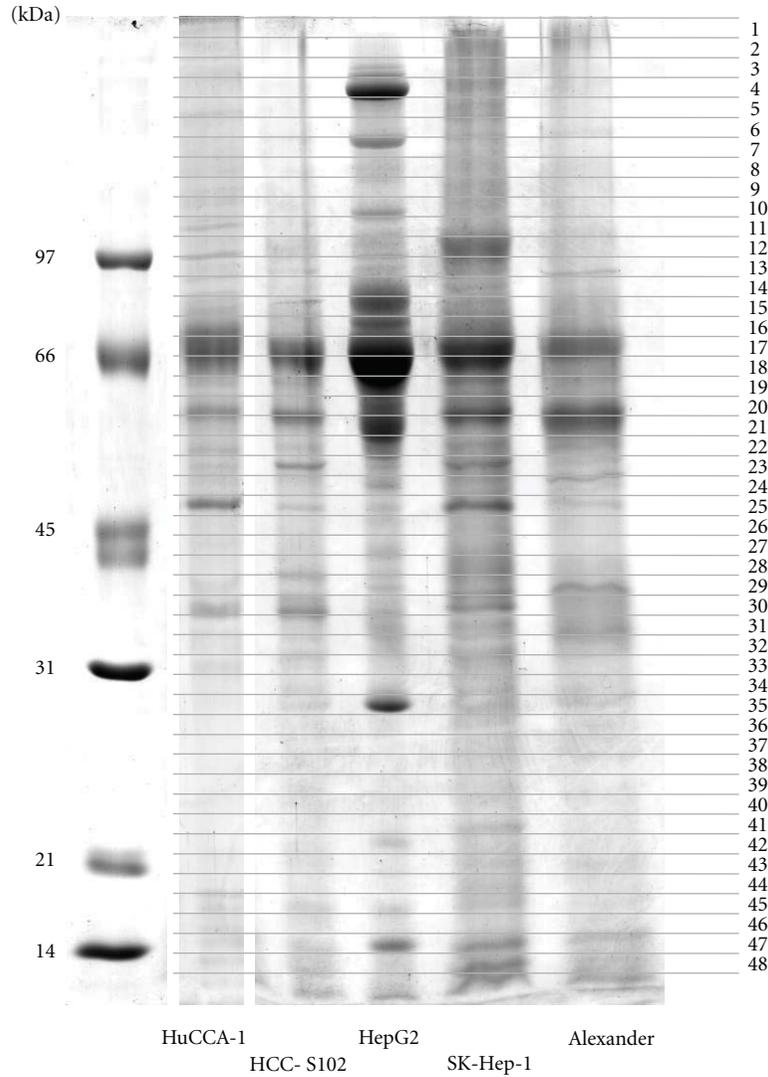
FIGURE 1: SDS-PAGE analysis of secretomes obtained from five cell lines. HuCCA-1, HCC-S102, HepG2, SK-Hep-1, and Alexander are shown in order from left to right, with MW standards in the leftmost lane. Each gel was cut into 0.5 mm slices, numbered as shown from the top of the gel to the dye front. Proteins in each slice were reduced, alkylated, digested with trypsin, and subjected to LC/MS/MS analysis.
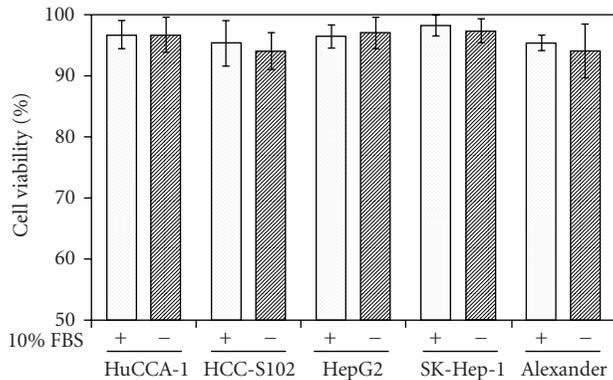
FIGURE 2: Viability of cancer cells cultured in complete media or serum-free media for 24 hours. Cell viability was determined by trypan blue exclusion. Data are expressed as mean ± SD of from three independent experiments.

oral). Thus, some 30%–50% proteins were expressed at the same level in 5 cell lines, while 50%–70% proteins showed differences in expression. Forty-nine secreted proteins were found only in HuCCA-1 which was marked in Table 2, including laminin 5 beta 3, heat shock 90 kDa protein 1, heat shock 70 kDs protein 8 isoform 1 variant, GRP78 precursor, desmoglein-4 nephroblastoma overexpressed precursor, neutrophil gelatinase-associated lipocalin (lipocalin 2, NGAL), desmoplakin, cathepsin D, DnaJ homolog subfamily B member 11, annexin IV variant, annexin A5, Ras-related protein Rap-1A, RHOV protein, and rotatin isoform CRA_e.

4. Validation of Potential Biomarkers

Of the overexpressed proteins found only in HuCCA-1, 15 appeared to be proteins related to cancer. 1-DE immunoblotting was used to verify the presence of 4 proteins by

TABLE 2: Identified secreted proteins from CCA cell line (HuCCA-1).

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
1	TSP1_HUMAN	Thrombospondin-1		129.30/ 4.71	77	7	7	K.GPDPSSPAFR.I K.FQDLVDAVR.A K.GGVNDNFQGVLQNVR.F	Cytoskeleton/ mobility
2	K2C6B_HUMAN	Keratin, type II cytoskeletal 6B	√	59.96/ 8.05	122	4	8	R.FVFGTTPEDILR.N R.FTGSQPFGQGVEHATAN K.Q K.GTSQNDPNWVVR.H R.AQGYSGLSVK.V R.SGFSSISVSR.S K.FASFIDKVR.F K.YEELQVTAGR.H R.ATGGGLSSVGGGSS TIK.Y	Cytoskeleton/ mobility
3	MYH10_HUMAN	Myosin-10		228.80/ 5.44	35	2	0	K.ESLTK.L K.NDNSSR.F	Cytoskeleton/ mobility
4	LAMC1_HUMAN	Laminin subunit gamma-1		177.49/ 4.82	42	1	4	K.TEQQTADQLLAR.A	Cytoskeleton/ mobility
5	MYH10_HUMAN	Myosin-10		228.80/ 5.44	37	3	0	K.NDNSSR.F K.ESLTK.L R.GCLARK.A R.AEDEAYYYCSSYAD	Cytoskeleton/ mobility
5	gi 4761370	Immunoglobulin lambda light chain variable region	√	10.05/ 9.17	36	2	22	TKCFGVR.R K.AFGVR.R K.LEAVGTGIEPKAMS	Immunological response
5	ZNF28_HUMAN	Zinc finger protein 28	√	98.64/ 9.44	38	2	5	QGLVTFGDVAVDFSQ EEWEWLNPIQR.N K.NFQKSSVVIK.Q	DNA replication /gene regulation
6	STK10_HUMAN	Serine/threonine-protein kinase 10	√	112.07/ 6.52	36	1	1	R.ELVAEAKAEVMEEED GR.D	Metabolism
6	PTPRH_HUMAN	Receptor-type tyrosine-protein phosphatase H	√	122.28/ 5.21	35	1	0	R.TETRNTTDTR.V	Metabolism
6	Q9HC04_HUMAN	Cytochrome P450 monooxygenase	√	49.50/ 5.98	38	1	2	K.TLDPFETMLK.S	Metabolism
6	MYST4_HUMAN	Histone acetyltransferase MYST4	√	199.70/ 5.05	39	2	2	K.KPEELLSCADCGSSGH PSCLKFCPELTTNVK.A R.ANSRQSPAK.V	DNA replication/ gene regulation

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
	PTPRF_HUMAN	Receptor-type tyrosine-protein phosphatase F		213.83/5.96	37	2	1	R.YSAPANLYVR.V R.AAGTEGPFQEVDGV ATTR.Y R.ELTDLNQEFETLQEK.A K.EINSLQSDFTK.Y	Transport/binding proteins
7	gi 119621568	HCG1811249, isoform CRA_a	√	149.08/7.76	59	5	4	R.EAELQVDQILTK.S R.LLIDDQLLR.N R.LSLSPEVLDLTSNSLK.R	Unknown
	ITA2_HUMAN	Integrin alpha-2	√	129.22/5.10	38	3	2	R.FGIAVLGYLNR.N K.IPLLYDAEIHLTR.S K.IGQTSSSVSFK.S	Transport/binding proteins
8	LAMB3_HUMAN	Laminin-5 beta 3	√	129.33/7.30	36	2	1	K.IQEVGEITNLR.V R.LPNVDLVLSQTK.Q	Transport/binding proteins
9	ITA6_HUMAN	Integrin alpha-6	√	118.82/6.36	36	1	1	K.DEITFVSGAPR.A	Transport/binding proteins
10	RAP2B_HUMAN	Ras-related protein Rap-2b	√	20.49/4.52	40	1	14	K.EIEVDSSPSVLEILDTAG TEQFASMR.D	Signal transduction
	RAP2C_HUMAN	Ras-related protein Rap-2c	√	20.73/4.67	40	1	14		Signal transduction
	RAP2A_HUMAN	Ras-related protein Rap-2a	√	20.60/4.52	40	1	14		Signal transduction
11	EF2_HUMAN	Elongation factor 2		95.28/6.39	36	3	3	R.VFSGLVSTGLK.V M.VNFTVDQIR.A K.YEWDVAEAR.K R.DLLLDPAWEK.Q K.LASDLLEWIR.R K.ASIHEAWTDGK.E	Protein synthesis and degradation
	ACTN4_HUMAN	Alpha-actinin-4		102.20/5.27	87	7	9	R.EAILAIHK.E R.QFASQANVVGPWIQ TK.M K.GISQEQMQEFR.A R.ETTDTDTADQVIASFK.V	Transport/binding proteins
12	DSG4_HUMAN	Desmoglein-4	√	113.75/4.42	43	1	2	R.EGIKGSSLLNYVLGTY TAIDLDTGNPATDVR.Y	Transport/binding proteins

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
13	gi 153792590	Heat shock 90 kDa protein 1	√	98.10/5.07	97	6	10	R.ELISNSSDALDK.I K.EDQTEYLEER.R K.YIDQEELNK.T K.TKPIWTR.N K.HIYYITGETK.D K.DQVANSAFVER.L	Chaperone/stress response
14	EZRI_HUMAN	Ezrin		69.37/5.94	42	4	6	K.APDFVFYAPR.L K.IALLEEAR.R R.QLLTLSSELSQAR.D K.IGFPWSEIR.N R.ITPSYVAFTPEGER.L	Cytoskeleton/mobility
	gi 386758	GRP78 precursor	√	72.29/5.07	61	4	7	K.TFAPEEISAMVLTK.M R.IINEPTAAAIAYGLDK.R R.ITPEEIER.M	Chaperone/stress response
15	TRFL_HUMAN	Lactotransferrin		78.13/8.01	35	1	2	R.YYGYTGAFR.C	Transport/binding proteins
	MOES_HUMAN	Moesin		67.78/6.08	43	2	3	R.EDAVLEYLK.I K.APDFVFYAPR.L	Cytoskeleton/mobility
16	Q53GZ6_HUMAN	Heat shock 70 kDa protein 8 isoform 1 variant	√	70.85/5.37	96	5	8	R.TTPSYVAFTDTER.L K.SFYPEEVSSMVLTK.M K.DAGTIAGLNVLR.I R.FEELNADLFR.G K.LLQDFFNGK.E	Chaperone/stress response
17	ALBU_HUMAN	Human serum albumin		69.32/5.86	61	3	3	K.VPQVSTPTLVEVSR.N R.LHCASLQK.F K.KVPQVSTPTLVEVSR.N	Transport/binding proteins
18	Q16503_HUMAN	Thyrotropin receptor-I	√	28.44/6.95	25	1	2	K.ELPLLK.F	Signal transduction

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
19	KPYM_HUMAN	Pyruvate kinase isozymes M1/M2		57.88/ 7.95	119	7	13	K.GSGTAEVELKK.G K.ITLDNAYMEK.C K.IYVDDGLISLQVK.Q K.GVNLPGAAVDLPA VSEK.D K.DIQDLK.F R.GDLGIEIPAEK.V K.VFLAQK.M	Metabolism
20	PDIA3_HUMAN	Protein disulfide-isomerase A3	√	56.75/ 5.92	36	1	2	R.ELSDFISYLQR.E	Metabolism
	PDIA1_HUMAN	Protein disulfide-isomerase		56.64/ 6.10	40	2	6	K.TFSHELSDFGLESTA GEIPVVAIR.T R.FLQDYFDGNLK.R	Protein synthesis and degradation
	HERC6_HUMAN	Probable E3 ubiquitin-protein ligase HERC6		46.39/ 7.19	39	1	2	K.ITLDEK.K	Metabolism
	RBM27_HUMAN	RNA-binding protein 27	√	44.78/ 9.54	41	1	3	K.TSSAVSTPSKVK.T	DNA replication/ gene regulation
21	K2C7_HUMAN	Keratin, type II cytoskeletal 7		51.39/ 5.42	45	2	5	R.EVTINQSLLAPLR.L R.LPDIFEAQIAGLR.G R.TMSEVGGSVEDLIAK.G K.QPGITFIAAK.F	Cytoskeleton/ mobility
	CATD_HUMAN	Cathepsin D	√	44.52/ 6.09	77	6	16	K.FDGILGMAYPR.I K.LVDQNIFSFYLSR.D K.VSTLPAITLK.L K.LSPEDYTLK.V	Protein synthesis and degradation
22	CATD_HUMAN	Cathepsin D	√	44.52/ 6.09	88	4	9	K.LVDQNIFSFYLSR.D K.QPGITFIAAK.F R.YYTVFDR.D R.VGFAEAAR.L	Protein synthesis and degradation
	EF1A1_HUMAN	Elongation factor 1-alpha 1		47.43/ 9.00	49	2	5	R.EHALLAYTLGVK.Q K.IGGIGTVPVGR.V	Protein synthesis and degradation
	TBB5_HUMAN	Tubulin beta chain		49.73/ 4.75	43	2	4	R.ISVYYNEATGGK.Y R.FPGQLNADLR.K	Cytoskeleton/ mobility

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
23	ENOA_HUMAN	Alpha-enolase		47.14/7.17	100	6	15	R.AAVPSGASTGIYEA LELR.D / K.KLNVTEQEK.I / K.LMIEMDGTENK.S / K.GVPLYR.H / K.LAMQEFMILPVGAA NFR.E / K.YDLDFK.S	Metabolism
	gi 119590453	EDAR-associated death domain, isoform CRA_a	√	42.32/5.64	42	2	5	K.KLNVTEQEK.I / K.LMIEMDGTENK.S	Unknown
24	ACTB_HUMAN	Actin, cytoplasmic 1		41.71/5.14	107	3	9	K.AGFAGDDAPR.A / R.GYSFTTTAER.E / K.SYELPDGQVITIGNER.F	Cytoskeleton/mobility
	K1C18_HUMAN	Keratin, type I cytoskeletal 18		48.03/5.16	40	1	3	R.AQIFANTVDNAR.I	Cytoskeleton/mobility
25	ACTS_HUMAN	Actin, alpha skeletal muscle	√	42.02/5.23	123	7	14	K.AGFAGDDAPR.A / R.AVFPSIVGRPR.H / K.RGILTLK.Y / R.GILTLK.Y / K.IWHHTFYNELR.V / R.LDLAGR.D / K.EITALAPSTMK.I	
	ACTB_HUMAN	Actin, cytoplasmic 1		41.71/5.29	162	7	22	K.AGFAGDDAPR.A / R.AVFPSIVGRPR.H / K.RGILTLK.Y / K.IWHHTFYNELR.V / R.VAPEEHPVLLTEAP LNPK.A / R.LDLAGR.D / R.GYSFTTTAER.E / K.EITALAPSTMK.I	Cytoskeleton/mobility
26	SH2D7_HUMAN	SH2 domain-containing protein 7	√	49.78/5.96	35	1	1	R.MNQAR.L	Transport/binding proteins
	K2C8_HUMAN	Keratin, type II cytoskeletal 8		30.84/4.91	37	1	4	R.LEGLTDEINFLR.Q	Cytoskeleton/mobility
27	DJB11_HUMAN	DnaJ homolog subfamily B member 11	√	40.49/5.74	36	1	3	R.TLEVEIEPGVR.D	Chaperone/stress response

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
28	gi 10567590	Sodium bicarbonate cotransporter-like protein	✓	118.58/ 8.00	45	2	4	R.LCHAQSRSMNDISLT PNTDQR.K R.LLGNSCKFIPDLALM SFILFFGTYSMTLTLK.K M.PYQYPALTPEQK.K K.ELSDIAHR.I	Transport/ binding proteins
	ALDOA_HUMAN	Fructose-bisphosphate aldolase A		39.40/ 8.30	52	5	20	K.GILAADESTGSIAK.R R.QLLTADDR.V R.TVPPAVTGITFLSGG QSEEEASINLNAINK.C	Metabolism
29	AFF4_HUMAN	AF4/FMR2 family member 4		69.38/ 9.45	38	1	1	K.NSSSTSKQK.K	DNA replication/ gene regulation
	ANXA2_HUMAN	Annexin A2		38.55/ 7.57	72	7	24	R.DALNIETAIK.T K.GVDEVTIVNILTNR.S R.QDIAFAYQR.R K.SALSGHLETVILGLLK.T K.TPAQYDASELK.A R.TNQELQEINR.V K.DIISDTSGDFRK.L	Signal transduction
	ANXA1_HUMAN	Annexin A1		38.69/ 6.63	36	1	3	K.TPAQFDADELR.A	Signal transduction
	gi 182073	Erythroid protein 4.1 isoform A	✓	86.55/ 5.31	42	2	2	K.APIAAPEPELK.T K.TVKGGISETR.I	Metabolism
30	ANXA2_HUMAN	Annexin A2		38.55 /7.57	70	7	21	K.AYTNFDAER.D R.DALNIETAIK.T K.GVDEVTIVNILTNR.S R.QDIAFAYQR.R K.TPAQYDASELK.A R.TNQELQEINR.V K.SLYYIQQDTK.G K.GVDEATIIDILTK.R K.TPAQFDADELR.A	Signal transduction
	ANXA1_HUMAN	Annexin A1		38.69/ 6.63	52	4	15	R.SEDFGVNEDLADSDAR.A	Signal transduction
	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase		36.03/ 8.26	36	2	6	K.GTDVNVFNTILTTR.S K.VGVNGFGR.I R.GALQNIIPASTGAAK.A	Metabolism
	K1C9_HUMAN	Keratin, type I cytoskeletal 9		62.09/ 5.19	35	2	4	R.SGGGGGGGLGSGGSIR.S K.STMQELNSR.L	Cytoskeleton/ mobility

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
31	ANXA1_HUMAN	Annexin A1		38.69/ 6.63	47	3	14	K.GVDEATIIDILTK.R K.GGPGSAVSPYPTFNP SSDVAALHK.A K.TPAQFDADELR.A	Signal transduction
	ANXA5_HUMAN	Annexin A5	√	35.91/ 4.73	35	1	4	R.GTVTDFPGFDER.A	Signal transduction
32	HRX_HUMAN	Histone-lysine N-methyltransferase HRX	√	431.68/ 9.23	36	3	0	R.GRPPSTERIK.T R.EPTFR.W R.GPRIK.H	DNA replication/ gene regulation
33	A8K1N0_HUMAN	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide, isoform CRA_a	√	27.73/ 4.73	77	5	22	K.SVTEQGAELSNEER.N R.NLLSVAYK.N K.VFYLK.M K.GIVDQSQQAYQ EAFEISK.K K.DSTLIMQLLR.D	Metabolism
34	gi 119610955	TSR1 20S rRNA accumulation homolog yeast isoform CRA b	√	50.28/ 9.11	38	1	2	K.LLLLDTQQEAGMLLR.Q	DNA replication/ gene regulation
	K1C18_HUMAN	Keratin, type I cytoskeletal 18		47.31/ 5.27	40	2	4	R.AQIFANTVDNAR.I R.IVLQIDNAR.L	Cytoskeleton/ mobility
35	TPIS_HUMAN	Triosephosphate isomerase		26.65/ 6.50	47	2	11	R.HVFGESDELIGQK.V K.VVLAYEPVWAIGTGK.T	Metabolism
36	FLNA_HUMAN	Filamin-A isoform CRA_e	√	263.75/ 5.79	35	2	0	K.DGTVTVR.Y K.RDSLPR.L M.ECYIVISSTHLSNG HFRNIK.G R.SLVLQNDMK.H	Cytoskeleton/ mobility
	Z804A_HUMAN	Zinc finger protein 804A	√	136.80/ 8.19	49	2	2		DNA replication/ gene regulation
	A0N4V7_HUMAN	HCG2039797	√	2.21/ 9.79	36	1	38	K.GITLSVRP.-	Unknown
	gi 119569964	HCG2040487	√	21.65/ 8.60	39	1	10	R.IQGKDLTVWTTQDV EGILGAK.G	Unknown
37	PRDX1_HUMAN	Peroxiredoxin-1		22.10/ 8.27	45	3	15	R.TIAQDYGVLK.A R.QITVNDLPVGR.S R.LVQAFQFTDK.H	Protection and detoxification
	gi 16075531	Immunoglobulin kappa chain variable region	√	9.11/ 9.13	36	1	13	R.SGTDFTLKISR.V	Immunological response

Table 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
	NOV_HUMAN	Nephroblastoma overexpressed precursor	√	39.14/ 8.12	39	2	7	K.QTRLCMVRPCEQEP EQPTDK.K K.AIHLQFK.N	DNA replication/ gene regulation
38	NGAL_HUMAN	Neutrophil gelatinase-associated lipocalin, NGAL, Lipocalin 2	√	20.54/ 8.87	38	1	6	K.SYPGLTSYLVR.V	Transport/ binding proteins
	PRDX1_HUMAN	Peroxiredoxin-1		22.10/ 8.27	39	3	15	R.TIAQDYGVLK.A R.QTVNDLPVGR.S R.LVQAFQFTDK.H	Protection and detoxification
39	DESP_HUMAN	Desmoplakin	√	201.24/ 6.68	40	2	1	K.EISMQKEDDSK.N R.GIVDSITGQR.L	Transport/ binding proteins
	NGAL_HUMAN	Neutrophil gelatinase-associated lipocalin, NGAL, Lipocalin 2	√	20.54/ 8.87	42	2	11	K.MYATIYELK.E K.SYPGLTSYLVR.V	Transport/ binding proteins
40	Q59FK3_HUMAN	Annexin IV variant	√	25.41/ 6.03	41	2	6	-.AASISR.L K.GDTSGDYR.K	Signal transduction
	FAM3C_HUMAN	Protein FAM3C		24.67/ 8.52	37	1	7	R.LIADLGSTSITNLGFR.D	Signal transduction
41	K1C16_HUMAN	Keratin, type I cytoskeletal 16		51.24/ 4.79	74	6	15	R.APSTYGGGLSVSSR.F R.LASYLDK.V R.ALEEANADLEVK.I K.IIAATIENAQPILQID NAR.L K.ASLENSLEETK.G R.LEQEIATYR.R R.SGGGGGGGLGSGGSIR.S	Cytoskeleton/ mobility
	K1C9_HUMAN	Keratin, type I cytoskeletal 9		62.09/ 5.19	68	5	8	R.LASYLDK.V K.TLLDIDNTR.M R.QGVDADINGLR.Q R.QEYEQLIAK.N	Cytoskeleton/ mobility
42	A8KAH9_HUMAN	Ras-related protein Rap-1A	√	20.97/ 6.38	42	2	14	K.SALTVQFVQGIFVEK.Y K.INVNEIFYDLVR.Q	Signal transduction

TABLE 2: Continued.

Bands	Accession numbers	Names	Proteins were not expressed in HCC cell lines	MW/pI	Score	Peptide matches	%Coverage	Matched sequences	Functions
	gi 150023192	NADH dehydrogenase subunit 5	✓	67.02/9.04	41	1	1	R.KMVGLLK.T	Metabolism
43	gi 62906706	Diacylglycerol kinase alpha	✓	28.08/4.98	35	1	2	K.KVSDVLK.L	Metabolism
	LIPH_HUMAN	Lipase member H	✓	50.83/7.14	35	1	1	R.KVAMVLK.E	Metabolism
	gi 119586924	Rotatin isoform CRA_e	✓	44.01/9.19	36	1	1	K.KEDGVIK.E	Transport/binding proteins
44	gi|158430561	Chain A, Crystal Structure Of The Bard1 Brct Repeat	✓	24.10/8.62	37	1	3	K.YEIPEGPR.R	DNA replication/gene regulation
45	PRDX5_HUMAN	Peroxiredoxin-5, mitochondrial		22.01/8.75	38	1	5	R.LLADPTGAFGK.E	Protection and detoxification
	Q05D03_HUMAN	RHOV protein	✓	27.79/8.93	40	2	7	R.AMPPR.E R.ACCYLECSALTQK.N	Transport/binding proteins
46	PROF1_HUMAN	Profilin-1		15.04/8.45	78	4	41	K.TFVNITPAEVGVLVGK.D R.SSFYVNGLTLGGQK.C R.DSLLQDGEFSMDLR.T K.STGGAPTFNVTVTK.T	Cytoskeleton/mobility
47	HBB_HUMAN	Hemoglobin subunit beta		15.99/6.87	39	2	18	R.LLVVYPWTKR.F K.LHVDPENFR.-	Transport/binding proteins
48	gi 119619136	HCG1790904, isoform CRA_b		37.27/9.67	40	2	4	K.SSILYRHK.V K.NEMLTFK.A	Unknown
	CENPF_HUMAN	Centromere protein F	✓	22.41/7.11	35	1	0	K.YTALEQKLK.K	Cell cycle

using specific antibodies for nephroblastoma overexpressed precursor and neutrophil gelatinase-associated lipocalin (lipocalin 2, NGAL). Immunodetection only confirmed the presence of lipocalin 2 (Figure 3(b)), but not the presence of the other two proteins. Expression of lipocalin 2 was found only in the conditioned media of HuCCA-1, but not in the conditioned media of the other 4 cell lines (Figure 3(b)). Furthermore, there was no expression of lipocalin 2 in the cell lysate of any of the 5 cell lines, including HuCCA-1 (Figure 3(a)).

The 2-DE proteomic pattern of conditioned media of HuCCA-1 with Coomassie blue staining is shown in Figure 4(a) and 2-DE immunoblotting of the same sample was used to detect the lipocalin 2 forms (Figure 4(b)), revealing 4 spots with MW/pI 23.5/6.0, 23.5/6.5, 23.0/7.3, and 22.5/8.7. These different forms are likely to result from posttranslational modifications.

To confirm that lipocalin 2 is highly expressed in cholangiocarcinoma, the 12 homogenate samples from the pairs of normal and cancer of cholangiocarcinoma tissues described in Table 1 were resolved by SDS-PAGE gels and immunoblotted with monoclonal antibody to lipocalin 2. The results clearly showed a positive band of 23 kDa in all cancer tissues (Figure 5), with 9 out of the 12 cases showing high expression level of lipocalin 2, while the corresponding normal tissues were all negative.

5. Discussion

The study of the cancer secretomes by using proteomic technology has greatly accelerated in recent years. With the rapid progress in mass spectrometry (MS), bioinformatics, and analytical techniques, cancer biomarker discovery has been greatly promoted by this approach. Our present work has investigated the cell secretomes of cholangiocarcinoma (CCA) compared to 4 hepatocellular carcinoma (HCC) cell lines. With each cell line, three repeat experiments were performed by collection of the conditioned media from CCA and HCC cell lines, followed by concentration and precipitation by TCA/acetone. The proteins were then run onto SDS-PAGE, gel bands excised consecutively, and proteins analyzed by LC/MS/MS. Some bands showed low levels of matching, some bands showed blue color, and some were almost clear. The pattern on SDS-PAGE of secreted proteins and the total identified proteins were differed with each cell line. When the protein expression of the HuCCA-1 cell line is compared to the HCC cell lines, only 14, 14, 19, and 6 proteins from Alexander, HCC-S102, HepG2, and SK-Hep-1, respectively, matched with HuCCA-1 (data not shown).

Proteins secreted into conditioned medium may be tumor specific and can represent potential biomarkers that may circulate in the blood stream. Our results showed 49 distinct proteins, expressed only in HuCCA-1. Some interesting secreted proteins expressed only in cholangiocarcinoma include laminin 5 beta 3, heat shock 90 kDa protein 1, heat shock 70 kDs protein 8 isoform 1 variant, GRP78 precursor, desmoglein-4 nephroblastoma overexpressed precursor, neutrophil gelatinase-associated lipocalin (lipocalin 2,

NGAL), desmoplakin, cathepsin D, DnaJ homolog subfamily B member 11, annexin IV variant, annexin A5, Ras-related protein Rap-1A, RHOV protein, and rotatin isoform CRA_e.

Many of these proteins are related to cancer. For example, laminin-5 (LAMA3, LAMB3, and LAMC2) is a heterotrimeric glycosylated protein that belongs to th Ln family and is formed by $\alpha 3\beta 3\gamma 2$ chains assembled with disulfide bonds. Ln-5 is widely expressed in the human body but shows differential expression in metastatic and nonmetastatic hepatocellular carcinoma [21, 22]. Ln-5 also plays an important role in cell migration during tumor invasion and tissue remodeling [22], and laminin 5 $\gamma 2$ chain and $\beta 3$ chain have both been suggested to be important in the invasiveness of cancer cells. Unfortunately, no commercial antibody appears to be available for laminin 5 $\beta 3$ chain; expression of the protein could not be validated in cholangiocarcinoma tissues. Laminin $\gamma 2$ chain exhibits aberrant expression in a stepwise manner through different aggressive stages of tumor progression [23].

Nephroblastoma overexpressed (NOV) belongs to the CCN family of genes that encode secreted proteins associated with the extracellular matrix (ECM) and exert regulatory effects at the cellular level. NOV is likely to play a role in cell growth regulation, in the progression and in the metastatic potential of melanomas. The expression of this protein appears to be associated with a higher risk of developing metastases in Ewing's sarcoma [24]. We found the high expression of NOV secreted in the media of the cholangiocarcinoma cell line.

Neutrophil gelatinase-associated lipocalin (NGAL or lipocalin 2) is a prominent member of the lipocalin family. It was first isolated as a 25 kda glycoprotein covalently bound with matrix metalloproteinase-9 (MMP-9) in human neutrophil [25]. It is a secreted acute phase protein, which is also upregulated in multiple cancers, including breast, lung, and pancreas. Recently lipocalin 2 has been proposed as an early biomarker in pancreatic cancer [26].

Desmoplakin, a specialized adhesion junction protein and the principal plaque protein of desmosomes, has been found in various tissues including heart, skin, and meninges [27]. The biological significance of desmoplakin has been recently reported in both autosomal dominant and autosomal recessive disorders from naturally occurring human gene mutation. Desmoplakin was found to be abnormal in many conditions including autoimmune blistering diseases, epithelial malignancies, and blood vessel morphogenesis. Desmoplakin was also found to be secreted by the cholangiocarcinoma cell line in our study.

Cathepsin D, the aspartic protease, is an independent marker of poor prognosis in human breast cancer [28]. Cathepsin-D was reported to play an essential role in multiple tumor progression steps, affecting cell proliferation, angiogenesis, and apoptosis. Another report also suggested that cathepsin D was a key mediator in induced apoptosis [29]. Our results showed highly expression levels of cathepsin D in the cholangiocarcinoma cell line. A cathepsin D-like protein has also been reported from the gut and other tissues of the human liver fluke *Opisthorchis viverrini* where longstanding infection is associated with cancer of the bile ducts
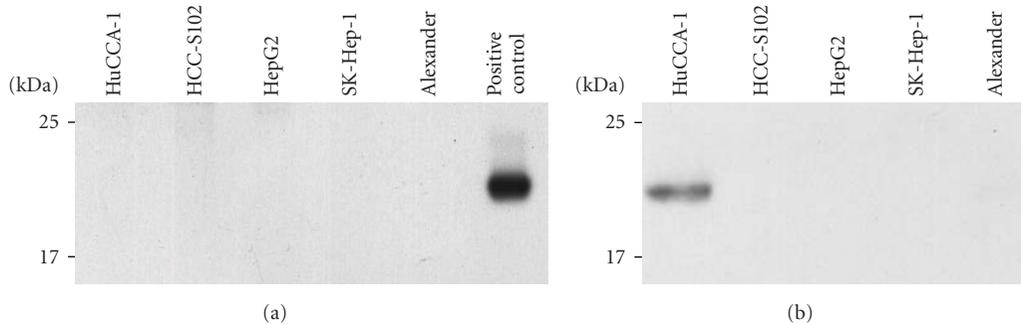
Figure 3: The western blots for immunoreactive lipocalin 2 from the lysate of five cell lines (HuCCA-1, HCC-S102, HepG2, SK-Hep-1, and Alexander) (a) and five conditioned media samples (b).
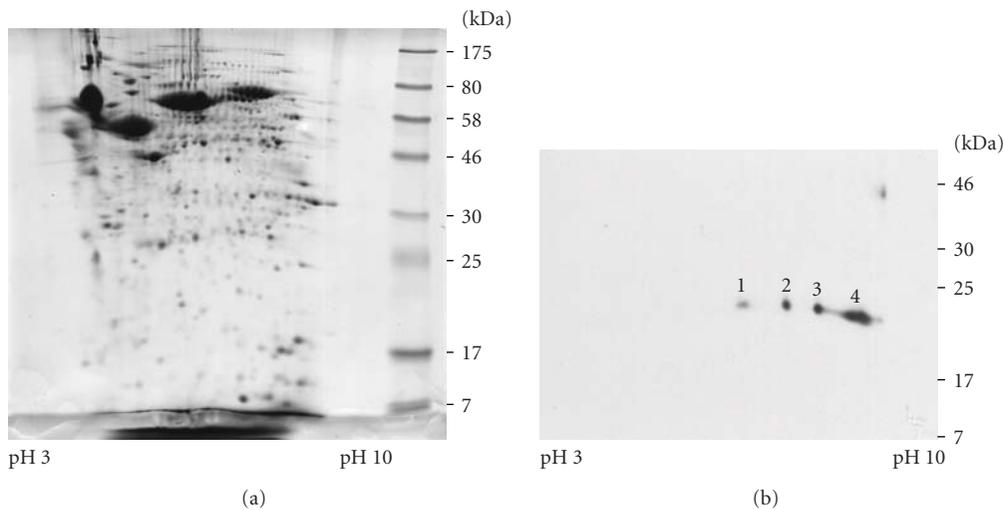

Figure 4: Two-dimensional gel electrophoresis of conditioned media from HuCCA-1 stained with Coomassie Blue R-250 (a) and immunoblot labeled with the antibody to lipocalin 2 (b). The pI and MW of spot numbers 1, 2, 3, and 4 were approximately 6.0/23.5, 6.5/23.5, 7.3/23, and 8.7/22.5 respectively.

or cholangiocarcinoma [30]. This protein was also found in the excretory and secretory products of cultured adult flukes, indicating a role in host-parasite relationships.

The annexins are family of calcium-regulated phospholipid-binding proteins with diverse roles in cell biology. There are 12 human annexin subfamilies (A1–A11 and A13) that have been found to have various intra- and extracellular roles in a range of cellular processes such as cell signalling, ion transport, cell division, and apoptosis [31, 32]. Individual annexins have been implicated in tumour development and progression. Annexins A1, A2, and A5 interact with cytoskeletal proteins. Annexin A1 also has a role in controlling the inflammatory response while annexin A2 is present on the external surface of endothelial cells. Annexin A2 and annexin A4 appeared to be potential markers of interest for diagnosis of colorectal cancer [33]. The present studies showed that annexin A1, annexin A2, annexin IV variant, and annexin A5 were expressed in the cholangiocarcinoma cell line.

The presence of the proteins nephroblastoma over-expressed precursor and neutrophil gelatinase-associated lipocalin (lipocalin 2, NGAL) was validated in the cell lysates and conditioned media from all 5 cells by using specific antibodies. Lipocalin 2 was found to be clearly expressed in only conditioned media. Then 2-DE immunoblot-ting of the conditioned media showed the expression of lipocalin 2 as 4 spots at MW/pI 23.5/6.0, 23.5/6.5, 23.0/7.3, and 22.5/8.7 (Figure 4(b)). The posttranslation modification of the lipocalin 2 has not been reported yet and further studies will be needed to provide more information.

Finally, we verified the expression of lipocalin 2 in 12 homogenate samples from the pairs of normal and cancer tissues from cholangiocarcinoma patients. The lipocalin 2 antibody clearly detected a band of 23 kDa in all cancer tissues, while the pairs of normal tissues were all negative. Of the cancer tissues, 9 out of 12 cases showed high expression levels of lipocalin 2. However, since the cases studied here were in the late stage (grade 3-4) of cancer, it is of interest to see if tissues at the early stages of cholangiocarcinoma show lipoclain 2 expression. Further studies will investigate the expression of lipocalin 2 in different types and stages of

N T N T N T N T N T N T N T N T N T N1 T1 N1 T1 N1 T1 (kDa)
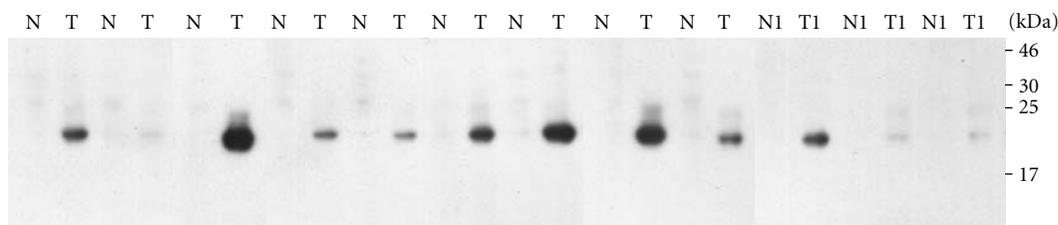
FIGURE 5: 1-DE patterns of the pairs of normal and cancer tissues from 12 cholangiocarcinoma patients. Details of the patients are shown from Table 1. The proteins were separated in 15% SDS-PAGE and detected using by immunoblotting with antibody to lipocalin 2.

cholangiocarcinoma in order to understand its relationship to grades and types of cancer.

6. Conclusion

Proteomic techniques have been used to compare the secretomes or proteins secreted into the external environment, cholangiocarcinoma, and hepatocellular carcinoma cell lines. Distinct proteins were found, in particular lipocalin 2, which may have potential value as a biomarker for cholangiocarcinoma.

Acknowledgments

References

[1] D. Sonakul, C. Koompirochana, K. Chinda, and T. Stitnimakarn, "Hepatic carcinoma with opisthorchiasis," *The Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 9, no. 2, pp. 215–219, 1978.

[2] S. B. Reddy and T. Patel, "Current approaches to the diagnosis and treatment of cholangiocarcinoma," *Current Gastroenterology Reports*, vol. 8, no. 1, pp. 30–37, 2006.

[3] S. L. Ong, A. Sachdeva, G. Garcea, et al., "Elevation of carbohydrate antigen 19.9 in benign hepatobiliary conditions and its correlation with serum bilirubin concentration," *Digestive Diseases and Sciences*, vol. 53, no. 12, pp. 3213–3217, 2008.

[4] B. E. Van Beers, "Diagnosis of cholangiocarcinoma," *HPB*, vol. 10, no. 2, pp. 87–93, 2008.

[5] S. Sell and H. A. Dunsford, "Evidence for the stem cell origin of hepatocellular carcinoma and cholangiocarcinoma," *American Journal of Pathology*, vol. 134, no. 6, pp. 1347–1363, 1989.

[6] C. Srisomsap, P. Sawangareetrakul, P. Subhasitanont, et al., "Proteomic analysis of cholangiocarcinoma cell line," *Proteomics. Clinical Applications*, vol. 4, no. 4, pp. 1135–1144, 2004.

[7] C. Srisomsap, P. Subhasitanont, P. Sawangareetrakul, et al., "Comparison of membrane-associated proteins in human cholangiocarcinoma and hepatocellular carcinoma cell lines," *Proteomics. Clinical Applications*, vol. 1, no. 1, pp. 89–106, 2007.

[8] I. Ladunga, "Large-scale predictions of secretory proteins from mammalian genomic and EST sequences," *Current Opinion in Biotechnology*, vol. 11, no. 1, pp. 13–18, 2000.

[9] H. Zwickl, E. Traxler, S. Staettner, et al., "A novel technique to specifically analyze the secretome of cells and tissues," *Electrophoresis*, vol. 26, no. 14, pp. 2779–2785, 2005.

[10] X. Lou, T. Xiao, K. Zhao, et al., "Cathepsin D is secreted from M-BE cells: its potential role as a biomarker of lung cancer," *Journal of Proteome Research*, vol. 6, no. 3, pp. 1083–1092, 2007.

[11] C. N. Perera, H. S. Spalding, S. I. Mohammed, and I. G. Camarillo, "Identification of proteins secreted from leptin stimulated MCF-7 breast cancer cells: a dual proteomic approach," *Experimental Biology and Medicine*, vol. 233, no. 6, pp. 708–720, 2008.

[12] A. M. Mlynarek, R. L. Balys, J. Su, M. P. Hier, M. J. Black, and M. A. Alaoui-Jamali, "A cell proteomic approach for the detection of secretable biomarkers of invasiveness in oral squamous cell carcinoma," *Archives of Otolaryngology*, vol. 133, no. 9, pp. 910–918, 2007.

[13] F. Mbeunkui, O. Fodstad, and L. K. Pannell, "Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS," *Journal of Proteome Research*, vol. 5, no. 4, pp. 899–906, 2006.

[14] V. Kulasingam and E. P. Diamandis, "Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets," *Molecular & Cellular Proteomics*, vol. 6, no. 11, pp. 1997–2011, 2007.

[15] C.-C. Wu, K.-Y. Chien, N.-M. Tsang, et al., "Cancer cell-secreted proteomes as a basis for searching potential tumor markers: nasopharyngeal carcinoma as a model," *Proteomics. Clinical Applications*, vol. 5, no. 12, pp. 3173–3182, 2005.

[16] H.-Y. Wu, Y.-H. Chang, Y.-C. Chang, and P.-C. Liao, "Proteomics analysis of nasopharyngeal carcinoma cell secretome using a hollow fiber culture system and mass spectrometry," *Journal of Proteome Research*, vol. 8, no. 1, pp. 380–389, 2009.

[17] S. Sirisinha, T. Tengchaisri, S. Boonpucknavig, et al., "Identification and potential use of a soluble tumor antigen for the detection of liver-fluke-associated cholangiocarcinoma induced in a hamster model," *Asian Pacific Journal of Allergy & Immunology*, vol. 9, pp. 153–157, 1991.

[18] K. Laohathai and N. Bhamarapravati, "Culturing of human hepatocellular carcinoma. A simple and reproducible method," *American Journal of Pathology*, vol. 118, no. 2, pp. 203–208, 1985.

[19] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding," *Analytical Biochemistry*, vol. 72, no. 1-2, pp. 248–254, 1976.

[20] T. Zakarias, J. Bunkenborg, M. Gronborg, et al., "A proteomic analysis of human bile," *Molecular & Cellular Proteomics*, vol. 3, no. 7, pp. 715–728, 2004.

[21] G. Giannelli, E. Fransvea, C. Bergamini, F. Marinosci, and S. Antonaci, "Laminin-5 chains are expressed differentially in metastatic and nonmetastatic hepatocellular carcinoma," *Clinical Cancer Research*, vol. 9, no. 10, pp. 3684–3691, 2003.

[22] S. Akimoto, Y. Nakanishi, M. Sakamoto, Y. Kanai, and S. Hirohashi, "Laminin 5 β3 and γ 2 chains are frequently coexpressed in cancer cells," *Pathology International*, vol. 54, no. 9, pp. 688–692, 2004.

[23] S. Aishima, S. Matsuura, T. Terashi, et al., "Aberrant expression of laminin gamma 2 chain and its prognostic significance in intrahepatic cholangiocarcinoma according to growth morphology," *Modern Pathology*, vol. 17, no. 8, pp. 938–945, 2004.

[24] V. Vallacchi, M. Daniotti, F. Ratti, et al., "CCN3/nephroblastoma overexpressed matricellular protein regulates integrin expression, adhesion, and dissemination in melanoma," *Cancer Research*, vol. 68, no. 3, pp. 715–723, 2008.

[25] S. Triebel, J. Bläser, H. Reinke, and H. Tschesche, "A 25 kDa α2-microglobulin-related protein is a component of the 125 kDa form of human gelatinase," *FEBS Letters*, vol. 314, no. 3, pp. 386–388, 1992.

[26] Z. Tong, A. B. Kunnumakkara, H. Wang, et al., "Neutrophil gelatinase-associated lipocalin: a novel suppressor of invasion and angiogenesis in pancreatic cancer," *Cancer Research*, vol. 68, no. 15, pp. 6100–6108, 2008.

[27] J. E. Lai Cheong, V. Wessagowit, and J. A. McGrath, "Molecular abnormalities of the desmosomal protein desmoplakin in human disease," *Clinical and Experimental Dermatology*, vol. 30, no. 3, pp. 261–266, 2005.

[28] G. Berchem, M. Glondu, M. Gleizes, et al., "Cathepsin-D affects multiple tumor progression steps in vivo: proliferation, angiogenesis and apoptosis," *Oncogene*, vol. 21, no. 38, pp. 5951–5955, 2002.

[29] M. Beaujouin and E. Liaudet-Coopman, "Cathepsin D overexpressed by cancer cells can enhance apoptosis-dependent chemo-sensitivity independently of its catalytic activity," *Advances in Experimental Medicine and Biology*, vol. 617, pp. 453–461, 2008.

[30] S. Suttiprapa, J. Mulvenna, N. T. Huong, et al., "*Ov*-APR-1, an aspartic protease from the carcinogenic liver fluke, *Opisthorchis viverrini*: functional expression, immunolocalization and subsite specificity," *The International Journal of Biochemistry & Cell Biology*, vol. 41, no. 5, pp. 1148–1156, 2009.

[31] V. Gerke and S. E. Moss, "Annexins: from structure to function," *Physiological Reviews*, vol. 82, no. 2, pp. 331–371, 2002.

[32] V. Gerke, C. E. Creutz, and S. E. Moss, "Annexins: linking Ca^{2+} signalling to membrane dynamics," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 6, pp. 449–461, 2005.

[33] P. Alfonso, M. Cañamero, F. Fernández-Carbonié, A. Núñez, and J. I. Casal, "Proteome analysis of membrane fractions in colorectal carcinomas by using 2D-DIGE saturation labeling," *Journal of Proteome Research*, vol. 7, no. 10, pp. 4247–4255, 2008.

Research Article

Quantitative Proteomics Analysis of Maternal Plasma in Down Syndrome Pregnancies Using Isobaric Tagging Reagent (iTRAQ)

Varaprasad Kolla,[1] Paul Jenö,[2] Suzette Moes,[2] Sevgi Tercanli,[1] Olav Lapaire,[1] Mahesh Choolani,[3] and Sinuhe Hahn[1,3]

[1] *Department of Biomedicine, University Women's Hospital, 4031 Basel, Switzerland*
[2] *Mass Spectrometry, Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland*
[3] *Biomarker Discovery Laboratory, Department of Obstetrics and Gynecology, National University of Singapore, Singapore 19077*

Correspondence should be addressed to Sinuhe Hahn, shahn@uhbs.ch

Currently no specific biomarkers exist for the screening of pregnancies at risk for down syndrome (DS). Since a quantitative proteomic approach with isobaric labelling (iTRAQ) has recently been suggested to be highly suitable for the discovery of novel plasma biomarkers, we have now used this method to examine for potential quantitative changes in the plasma proteome of the pregnancies bearing DS fetuses in comparison to normal healthy babies. In our study, we used plasma from six women with DS pregnancies and six with uncomplicated pregnancies care were taken to match cases and controls for gestational and maternal age, as these could be a confounder. In our quantitative proteomics analysis we were able to detect 178 proteins using iTRAQ labelling in conjunction with 4800 MALDI TOF/TOF. Amongst these we observed changes in βHCG, a known screening marker for DS, indicating that our assay was functional. We found a number of elevated proteins Ig lambda chain C region, serum amyloid P-component, amyloid beta A4, and under expressed proteins like gamma-actin and titin in DS pregnancies. These proteins are also found in the sera of patients with Alzheimer disease, which share similar pathologies of DS. Our study therefore indicates that the iTRAQ labelling approach may be indeed useful for the detection of novel biomarkers.

1. Introduction

Down syndrome (DS) is the most common chromosomal aneuploidy in live births and a leading cause of mental retardation. Prenatal detection of chromosomal anomalies, such as trisomy 21, in cases with DS relies on invasive practices such as amniocentesis or chorionic villi sampling. These procedures are associated with a risk of fetal loss or adverse pregnancy outcome [1]. Furthermore, it requires special facilities and highly trained staff. For this reason alternative strategies are sought, such as the noninvasive prenatal diagnosis of fetal genetic anomalies via the analysis of rare trafficking fetal cells [2] or cell-free fetal nucleic acids in maternal blood [3]. As this long sought goal has not yet reached clinical implementation [4], current clinical practice relies on a series of screening steps aimed at detecting at-risk pregnancies. In many centres the screening procedure for

DS is carried out in the late first and early second trimester (11–14 weeks of gestation) and involves a combination of ultrasound and serum marker analysis [5]. Although the efficacy of this method has improved considerably, and is more accurate than previous second trimester screening approaches, it is hampered by the skill and precision of the ultrasonographer and external factors such as ethnicity or other factors, such as assisted reproductive technologies.

Hence, it is obvious that new tools are needed to improve current state of the art. One such strategy is to look for new potential biomarkers using noninvasive proteomic approaches. The rationale for this strategy is that placentae from fetuses with DS, especially the villi, are morphologically distinct from the placentae of euploid fetuses, a feature attributed to altered aberrant protein expression [6]. As the placenta, a large organ with rapid cell turn-over is in direct contact with the maternal circulation, proteins

TABLE 1: Patients with and without a down syndrome pregnancy at the time of screening $n = 6$.

	First trimester	
	DS	Control
Maternal age (years)	35.8 ± 4.3	35.1 ± 4.1
Gestational age (weeks)	12.4 ± 1.2	12.5 ± 1.1

released by it should be detectable in maternal plasma, and could consequently serve as promising markers for abnormal placentation involved in a number of pregnancy related disorders. In this regard, preliminary proteomic investigations have shown that quantitative alterations in protein expression are found to occur in the amniotic fluid [7, 8] and maternal serum of pregnancies with fetuses affected with down syndrome, and that these could serve as new potential biomarkers [9].

Unfortunately none of these studies used quantitative approaches of the type that permit precise assessment of the extent of up- or down-regulation of the proposed biomarkers. As such, it will be difficult to validate these in blinded studies of clinical serum or plasma samples, a crucial facet in determining their specificity [9–11].

These studies are further complicated by the complex nature of the plasma/serum proteome, whereby the majority of low abundance proteins are masked by the preponderance of a few highly abundant proteins. This high dynamic range [12] effectively precludes the use of more conventional proteomic strategies, such as those employing gel electrophoresis with or without or fluorescent labelling, for example, Difference Gel Electrophoresis (DIGE). This deficit renders the identification of potential biomarkers in plasma/serum by conventional comparative proteomic approaches very challenging.

For this reason an approach using Isobaric Tags for Absolute and Relative Quantitation (iTRAQ) has been proposed for the discovery of plasma biomarkers [13]. This chemical labelling method involves the stable incorporation of isotopes into an amine tagging reagent, which can be reliably detected by mass spectrometry, thereby permitting comparative quantitation in a multiplex manner. Currently 4-plex and 8-plex reagents are commercially available, which are used to label the protein samples of interest following trypsin digestion. The use of different isobaric tags implies that up to 4 or 8 different samples, one of which serves as a reference, can be examined simultaneously in a single mass spectrometric analysis. It is for this reason that the iTRAQ approach has been suggested to be suitable for the discovery of biomarkers in a wide range of body fluids and tissues, including plasma.

In order to examine whether this approach would be suitable for the detection of potential biomarkers for down syndrome, we performed a proof-of-principle experiment, in which we examined samples from 6 cases with down syndrome in comparison to 6 matching controls. In our study we have used a 4-plex iTRAQ labelling in conjunction with a 4800 MALDI TOF/TOF approach to examine plasma samples obtained in first trimester pregnancies. Our results indicate that quantitative differences can be detected between aneuploid samples and samples from euploid pregnancies and that such alteration may reflect upon changes known to occur in down syndrome.

2. Materials and Methods

2.1. Samples. Blood samples for this case-control proteome study were collected from six pregnant women carrying a DS fetus (11–13 weeks of gestation) and six pregnant women with normal euploid pregnancies. The samples were matched for maternal and gestational age (Table 1). This study was undertaken with the approval of the Institutional Ethical Board of the University Hospital, Basel, Switzerland and written informed consent was required in all instances.

2.2. Sample Preparation. 9 mL blood was drawn into BD P100 tubes (BD Diagnostics, Franklin Lake, NY, USA), which are specially designed for proteomics experiments, in that the Ethylenediaminetetraacetic acid (EDTA) and protease inhibitor present in the tube prevent coagulation and stabilize the plasma proteome. Following phlebotomy the samples were centrifuged at $3000 \times$ g for 30 minutes at $10°$C, whereby the plasma was separated from the cellular fraction by aid of a mechanical separator. $100 \mu L$ aliquots were stored at $-80°$C until further use. For an overview of the work-flow used in this analysis refer to Figure 1.

2.3. Immunodepletion of High-Abundance Plasma Proteins. Highly abundant plasma proteins were depleted using ProteoMiner Protein Enrichment Kit (Bio-Rad Laboratories, Inc.) [14], as per the manufacturer's instructions. 1 mL of plasma was used for the depletion and after the whole procedure, $300 \mu L$ was eluted in elution reagent. After depletion protein concentration was measured by using RC-DC Protein assay kit (Bio-Rad Laboratories, Inc.)

2.4. Tryptic Digestion and iTRAQ Reagent Labelling. Equal amounts $(100 \mu g)$ of depleted plasma protein from six of the DS cases and controls were pooled separately in duplicate for the iTRAQ labelling. These samples were denatured with 2% SDS in 500 mM triethylammonium bicarbonate (TEAB) (Sigma-Aldrich) for 15 minutes at room temperature, following which they were reduced with $2 \mu L$ of 50 mM tri-(2-carboxyethyl) phosphine (TCEP) (Sigma-Aldrich) at $60°$C for 1 hour and were then alkylated with 10 mM s-Methylmethanethisulfonate (MMTS) for 10 minutes in room temperature. After alkylation, the proteins were digested overnight at $37°$C with 1 U/μL trypsin (TPCK treated) (Applied Biosystems, Foster City, CA 94404, USA). Peptides were labelled with one unit of iTRAQ Reagent Multi-plex kit (Applied Biosystems, Foster City, CA 94404, USA) that was reconstituted in $70 \mu L$ of ethanol. iTRAQ labels 114, 116 were used separately for labelling the pooled duplicated control sample and 115, 117 were used separately to label the pooled duplicate down syndrome samples. The iTRAQ labelling reagent solution was added to the digest and incubated for 1 hour at room temperature. To assess the

Human plasma

↓

Depletion

↓

Reduced, cysteine block & trypsin digestion

↓

iTRAQ labelling

↓

Control: iTRAQ reagent (114)   DS: iTRAQ reagent (115)
Control: iTRAQ reagent (116)   DS: iTRAQ reagent (117)

↓

Combined labelled sample

↓

Strong cation exchange chromatography

↓

LC MS/MS for protein identification and quantitation
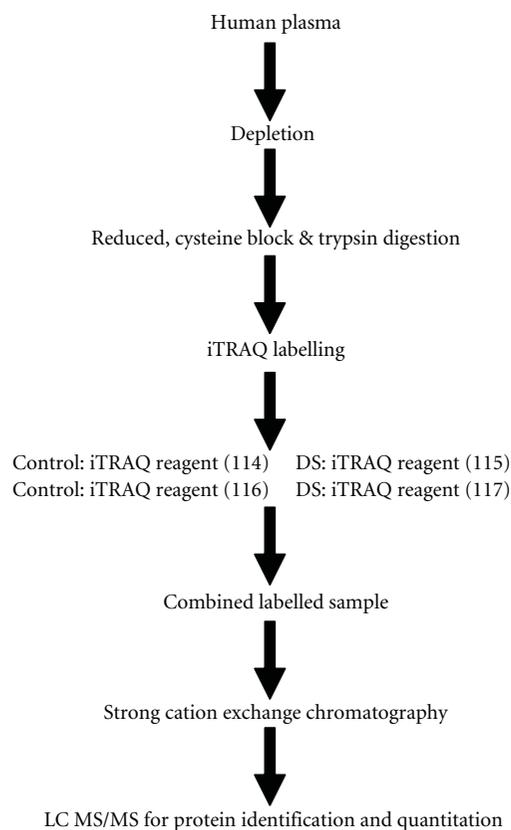
FIGURE 1: Workflow for quantitative proteomics using iTRAQ reagent. Equal amounts of plasma protein ($100\,\mu$g) from control and DS ($n = 6$) were pooled separately and duplicated, controls were labeled with 114 and 116 iTRAQ label and DS was labeled with 115 and 117 iTRAQ label. The labeled samples were pooled and were subjected to a strong cation exchange chromatography to remove the excess label. Afterwards LC-MALDI MS/MS was performed for protein identification and quantification.

accuracy of the ratiometric quantitation of iTRAQ reagent a split in signal was performed and the data were corrected as described in detail by Unwin and colleague [15].

2.5. Strong Cation Exchange Chromatography (SCX). Strong cation exchange chromatography was preformed to remove the excess iTRAQ reagent and interfering substances for the mass analysis. Dried peptides were resuspended in $200\,\mu L$ of Buffer A and were loaded on Poly Sulfoethyl A Column (200 mm length × 4.6 id, $5\,\mu$m particle size, 200 pore size) on a BioLC HPLC unit (Dionex). Buffer A consisted of 10 mM $KHPO_4$ and 25% acetonitrile, 500 mM KCl, pH 3.0, and Buffer B consisted of 10 mM KH_2PO_4, 25% acetonitrile, and 500 mM KCl pH 3.0. The 60 minutes gradient comprised of 100% Buffer A for 5 minutes, 5%–30% Buffer B for 40 minutes, 30%–100% Buffer B for 5 minutes, 100% Buffer B for 5 minutes, and finally 100% Buffer A for 5 minutes. Ten fractions were collected using a Foxy Jr. Fraction Collector (Dionex). Subsequently, these fractions were pooled according to the chromatogram profile based on the peak intensity and the products dried in a vacuum

concentrator, after which they were stored at $-20°$C prior to mass spectrometric analysis.

2.6. Nano LC MALDI. The dried SCX iTRAQ-labeled peptides were dissolved in Buffer A which consist of 95% H_2O, 5% acetonitrile, 0.1% TFA and were loaded on C18 trap column (1 mm × $300\,\mu$m i.d. column) at $30\,\mu L$/minutes and separated on an analytical column (150 mm ×$100\,\mu$m i.d. column) at 500 nL/min using the LC-packing Ultimate system. The peptides were separated using a linearly increasing concentration of acetonitrile in Buffer B from 5% to 30% in 120 minutes, and from 30% to 60% in 40 minutes. The elute was mixed with matrix (2 mg/mL alpha-cyano-4-hydroxycinnamic acid in 80% acetonitrile, and 0.1% TFA) at a flow rate of 800 nl/min and deposited on an Opti-TOF LC/MALDI (Applied Biosystems) plate in 10S fractions, using an automatic robot (Probot, Dionex).

2.7. MS and MSMS. The Mass spectrometer 4800 plus MALDI TOF/TOF Analyzer (Applied Biosystems) was set to perform data acquisition in positive ion mode. An MS condition of 1000 shots per spectrum was used. Monoisotopic precursor selection for MS/MS was done by automatic precursor selection using an interpretation method using the 12 most intense peaks per spot with an MS/MS mode condition of 4000 laser shots per spectrum. Minimum peak width was one fraction and mass tolerance was 80 ppm. Adduct tolerance is $(m/z) \pm 0.003$. MSMS was done with a gas pressure of 2×10^{-2} bar in the collision cell. Air was used as collision gas.

2.8. Protein Identification and Database Searches. Protein identification and quantification was done by using the ProteinPiolt software v2.0 (Applied Biosystems; MDS-Sciex). The search was performed against the Human database of UniProtKB/Swiss-Prot (Version 3.50) from the EBI website (http://www.ebi.ac.uk./IPI/IPIhelp.html) and concatenated target-decoy database search strategy was used to check the false positive rate [16] in our case it was found to be 0%, which boosted the reliability of our data.

2.9. Relative Quantitation Criteria. The Paragon algorithm [17, 18] in ProteinPiolt v2.0 software was used as the default search program with digestion enzyme set as trypsin and methyl methanethiosulfonate as cysteine modification. The search also included the possibility of more than eighty biological modifications and amino acid substitution of up to two substitutions per peptide using the BLOSUM 62 matrix. Data were normalized for loading error by bias correction calculated with Progroup algorithm Identified proteins with at least 95% confidence and with a ProtScore of 1.3, were reported. The results obtained from ProteinPiolt software v2.0 software were exported to Microsoft Excel for the further analysis. The study was performed in double duplex manner, where DS samples were labelled with 115 and 117, control were labelled with 114 and 116. Peptides were selected based on the criteria defined in the protein pilot software, which means all the peptides were included for quantitation with

an exception for those without an iTRAQ modification or reporter ion, an area count less than 40 and peptides with *P* value less than .001 were excluded [19]. As described by Gan and colleague in their study on estimation of relative quantitative ratio from iTRAQ experiments, we also used only peptides above or equal to 70% confidence level for the estimation of relative quantitation [20].

2.10. PANTHER Analysis. The PANTHER database was used to elucidate the molecular function, biological process and signaling pathway associated with each individual protein (http://panther.appliedbiosystems.com/).

3. Results

Samples were obtained from 6 cases with a confirmed DS fetus and 6 samples from normal euploid singleton pregnancies. Care was taken to match both maternal and gestational age, to rule out any confounding influence of these two parameters.

Low abundant plasma proteins were enriched by using ProteoMiner Protein Enrichment Kit. This was accomplished through the use of a large, highly diverse bead-based library of combinatorial peptide ligands. When plasma was applied to the beads, a small fraction of the high abundance proteins saturated their high affinity ligands and the excess high abundance proteins were washed away. In addition a very small amount of high abundance proteins and low abundance proteins were concentrated on their specific affinity ligands. This provides for a significant enrichment of medium and low abundant plasma proteins.

The samples were pooled separately in and duplicate in order to have more precise analytical replicate measurements. The iTRAQ analysis was done in double duplex style, the DS samples were labelled with iTRAQ 115 and 117 and the control samples with iTRAQ 114 and 116, using the work-flow illustrated in Figure 1.

Following tandem MS MS, and by focussing on iTRAQ reporter ions in low molecular mass range (114–117 Da) for quantification, we identified 235 proteins with ≥95% confidence. However, after manually rechecking the MS/MS data thoroughly peak by peak, only 187 out of 235 proteins (78.5%) had a relative quantitation derived from the analysis of two or more peptides, while for 45 proteins, the quantitation was based on single peptide. For 3 proteins no quantitation could be ascertained by analysis using Protein Pilot. Figure 2 shows the MSMS spectrum of the precursor ([M + H] +, *m/z* 1527.7 Da). In low-mass region the reporter ions, are seen while area under the curve was used for quantification.

As we did the experiment in double duplex manner, DS (115 and 117) and control (114 and 116), it was possible to estimate the cutoff point for differentially expressed protein in our sample [19, 20]. Based on 187 relative abundant protein ratios from DS and control sample, an average variation of 4.4% (±0.04) was measured. If the cutoff was set at 5% average variance then only 72% of the proteins would fall with in this variation range, but. If the range was


MSMS 1527.7 Da HQDGSYSTFGER
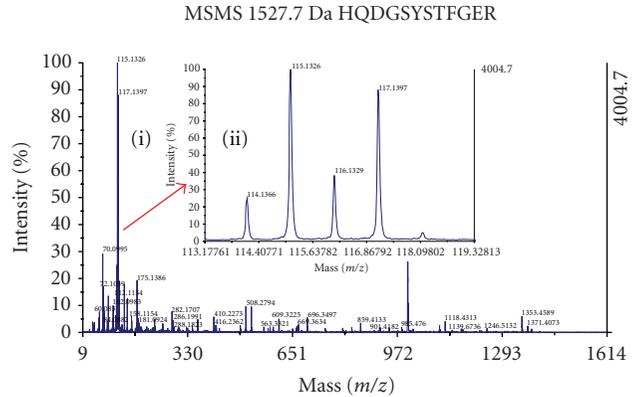
FIGURE 2: Components of the spectrum illustrated are (i) MSMS spectrum of the precursor ([M + H] +, *m/z* 1527.7 Da). (ii) low-mass region showing the reporter ions used for quantitation. The peptide is labeled by isobaric tags at both the N terminus and C-terminal lysine side chain. The precursor ion and all the fragment ions therefore contain all four members of the tag set, but remain isobaric. The MSMS spectrum was obtained from the singly charged [M + H] + peptide using a 4800 MALDI TOF-TOF analyzer.
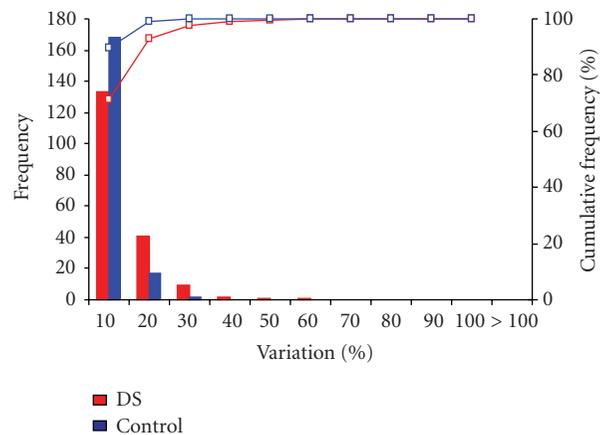

FIGURE 3: Frequency distribution (bars) from both DS and control replicates across different ranges of variation. The cumulative percentage (lines) is defined as the cumulative number of proteins falling within the defined variation range against the total number of protein.

increased to 20% then about 92% of the protein falls in this variation range as illustrated in Figure 3. So the cutoff point in this experiment was set at 20% (±0.2). That means any relative change in protein ratio below or above ±1.2 fold was considered as differentially under or over expressed.

The functional distribution of these proteins is illustrated in Figure 4. For this interpretation, an analysis of 235 proteins was performed using the PANTHER classification system, which sorts the proteins into respective classes based on their biological process.

It is of interest that two of the major groups involve cell adhesion molecules (13%) and extra cellular matrix proteins (18%) and member of the protease family, factors that are known to be aberrant in down syndrome. Further more large

TABLE 2: List of the protein identified as up-regulated from the iTRAQ experiment, indicating is the biological process and molecular function of these proteins. (P value. 001).

No.	Acc. number	% Cov	Protein name	Biological processes	Molecular functions
1	P20742	31.5	Pregnancy zone protein	Ligand-mediated signaling	Other cytokine
2	P04003	30	C4b-binding protein alpha chain	Complement-mediated immunity	Complement component
3	P01842	68.6	Ig lambda chain C regions	Immunity	Immunoglobulin
4	P01233	26.1	Choriogonadotropin subunit beta	mRNA transcription	Other signaling molecule
5	P43251	14.1	Biotinidase	Vitamin metabolism	Other hydrolase
6	P01834	84	Ig kappa chain C region	Immunity	Immunoglobulin
7	P01859	42	Ig gamma-2 chain C region	Immunity	Immunoglobulin
8	P01215	6	Glycoprotein hormones alpha chain	Protein targeting and localization	SNARE protein
9	P80108	16.5	Phosphatidylinositol-glycan	Intracellular signaling cascade	Lipase
10	P01860	53.8	Ig gamma-3 chain C region	B-cell- and antibody-mediated immunity	Immunoglobulin
11	Q9UGM5	27.5	Fetuin-B	Intracellular signaling cascade	Lipase
12	P35858	26	Insulin-like growth factor	Cell adhesion	Receptor
13	P26927	22.9	Hepatocyte growth factor-like protein	Ligand-mediated signaling	Growth factor
14	P02751	49.4	Fibronectin	Extracellular matrix protein	Cell adhesion molecule
15	P02743	54.3	Serum amyloid P-component	Amino acid biosynthesis	Synthase
16	P00751	37.7	Complement factor B	Proteolysis	Serine protease
17	P01031	42.6	Complement C5	Complement-mediated immunity	Complement component
18	P07358	36.5	Complement component C8 beta chain	Complement-mediated immunity	Complement component
19	P02790	67.3	Hemopexin	Vitamin/cofactor transport	Carrier protein
20	P00734	61.3	Prothrombin	Blood clotting	Serine protease
21	Q14624	54.8	Inter-alpha-trypsin inhibitor heavy chain H4	Proteolysis	Serine protease inhibitor
22	O00213	10.8	Amyloid beta A4	Other neuronal activity	Other signaling molecule
23	P08603	61.7	Complement factor H	Complement-mediated immunity	Complement component
24	P00738	53	Haptoglobin	Neurotransmitter release	Vesicle coat protein

TABLE 2: Continued.

| No.	Acc. Number	% Cov	Protein name	Biological processes	Molecular functions
25	P22891	10.3	Vitamin K-dependent protein Z	Proteolysis	Serine protease
26	P22792	14.1	Carboxypeptidase N subunit 2	Cell surface receptor	Receptor
27	P07357	27.9	Complement component C8	Complement-mediated immunity	Complement component
28	P02741	8.9	C-reactive protein	Stress response	Defense and immunity
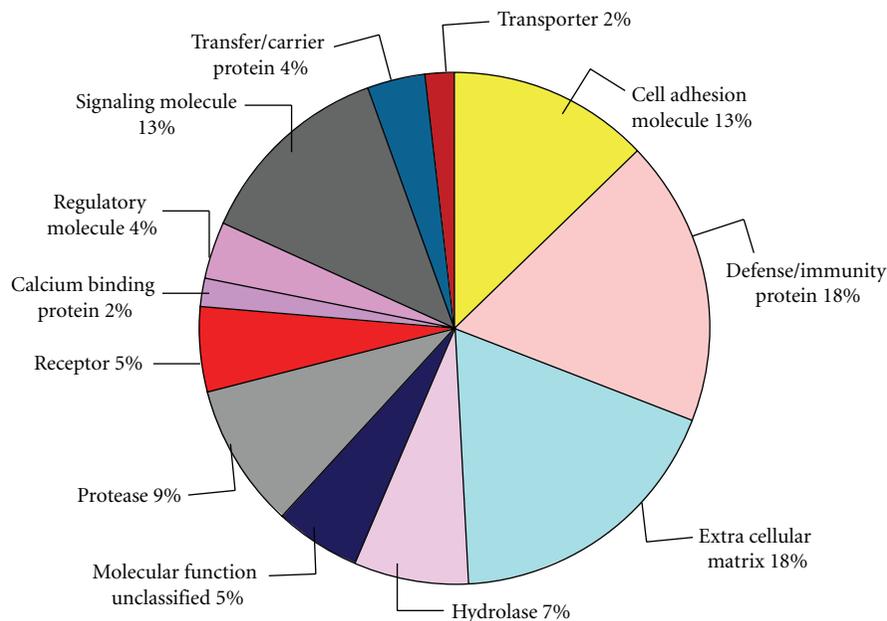
FIGURE 4: Number of plasma protein identified using iTRAQ reagent. In total, 235 proteins were identified. Shown above is the classification of these protein in different category based on molecular function.

groups were found to involve signalling molecules (13%) and 18% defense/immunity proteins.

Of the large group of significantly up-regulated proteins, it is highly noteworthy that this includes βHCG (Beta human chorionic gonadotropin), a subunit of the human chorionic gonadotropin protein known to be elevated in pregnancies [5] with DS fetuses, and which forms one of the backbone of current screening programs in the first trimester. This finding, therefore strongly suggests, that the iTRAQ method we have chosen appears to be able to identify relevant proteins (Table 2). In this group, elevations were also recorded for pregnancy zone protein (PZP) (P20742), thereby indicating that our assay is indeed detecting proteins of placental origin. Elevation of two members of the amyloid family (Serum amyloid P-component and Amyloid beta A4) was also observed, which could be a significant finding, as these proteins have been suggested to play a role in the dementia found to occur in adult DS patients. An elevation in the inflammatory response marker, C-reactive protein and members of the immunoglobulin family like Ig lambda chain C region was also noted, which could be indicative of an increased inflammatory response in DS pregnancies. phosphatidylinositol-glycan (P80108), Insulin-like growth factor (P35858) is involved in protein-protein interactions that result in protein complexes and hepatocyte growth factor (P26927) was also prominent amongst the list of the up-regulated proteins.

Amongst the group of down-regulated proteins (Table 3) were a number of molecules involved in cell adhesion and extracellular matrix including titin (Q8WZ42), basement membrane specific heparan sulfate proteoglycan (P98160), actin, cytoplasmic 2 (Gamma-actin) (P63261)and fibrinogen alpha chain (P02671). This observation could be important as changes in tissue elasticity are a hallmark of DS cases, and play significant role in their detection by ultrasound via the presence of an increased neck fold (nuchal translucency). peroxiredoxin 2 (PRDX2), an antioxidant enzyme, is under-expressed in DS fetal brain and dynein heavy chain 9, was also noted in our list of down regulated proteins.

The PANTHER database was used for the pathway analysis on the proteins which were shown to be over or under expressed in our study. It was interesting to note that

TABLE 3: List of the protein identified as down regulated from the iTRAQ experiment, indicating is the biological process and molecular function of these proteins. (*P* value .001).

No.	Acc. number	% Cov	Protein name	Biological processes	Molecular functions
1	P01009	32.8	Alpha-1-antitrypsin	Nerve-nerve synaptic transmission	Glutamate receptor
2	P02768	67	Serum albumin	Amino acid biosynthesis	Reductase
3	P63261	49.1	Actin, cytoplasmic 2 (Gamma-actin)	Cell structure	Actin and actin related protein
4	O95445	32.4	Apolipoprotein M (Apo-M)	Lipid transport	Plasma protein
5	P06727	67.7	Apolipoprotein A-IV	Lipid and fatty acid transport	Transporter
6	P02775	41.4	Platelet basic protein	Pyrimidine metabolism	Phosphorylase
7	P02647	85.8	Apolipoprotein A-I	Lipid and fatty acid transport	Transporter
8	P01023	39.8	Alpha-2-macroglobulin	Developmental processes	Serine/threonine kinase
9	P02787	58.9	Serotransferrin	Amino acid biosynthesis	Synthase
10	P04275	11.3	von Willebrand factor	Cell adhesion	Extracellular matrix glycoprotein
11	P02671	35.2	Fibrinogen alpha chain	Cell proliferation and differentiation	Extracellular matrix glycoprotein
12	P01024	29.6	Complement C3	Nerve-nerve synaptic transmission	Glutamate receptor
13	P98160	5.1	Heparan sulfate proteoglycan	Extracellular matrix	Cell adhesion mediated signal
14	P51884	33.4	Lumican	Receptor	Cell adhesion-mediated signaling
15	P04114	41.5	Apolipoprotein B-100	Lipid and fatty acid transport	Component of serum lipoproteins
16	Q8WZ42	9.5	Titin (EC 2.7.11.1) (Connectin)	Muscle contraction	Actin binding cytoskeletal protein
17	P08519	11	Apolipoprotein(a)	Cell proliferation and differentiation	Peptide hormone
18	P43652	47.7	Afamin	Transport	Other transfer/carrier protein
19	P06396	54.9	Gelsolin	Cell structure	Nonmotor actin binding protein
20	P02749	54.2	Beta-2-glycoprotein 1	Amino acid metabolism	Transaminase
21	Q9NYC9	4.8	Dynein heavy chain 9	Force generating protein	ATPase activity
22	P32119	4	Peroxiredoxin-2	Redox regulation	Eliminating peroxides

13.3% proteins identified in the DS sample correspond to proteins found in the Alzheimer disease-amyloid secretase and Alzheimer disease-presenilin pathways illustrated in Figure 5.This indicates that these pathways should be more closely studied in conjunction with DS. More than the 40% of the protein we identified in DS samples correspond to proteins in the Integrin signalling pathway. This might be due to the fact that cell adhesion molecules and extra cellular matrix protein represent 13% and 18% of the total proteins identified, respectively as illustrated in Figure 4.
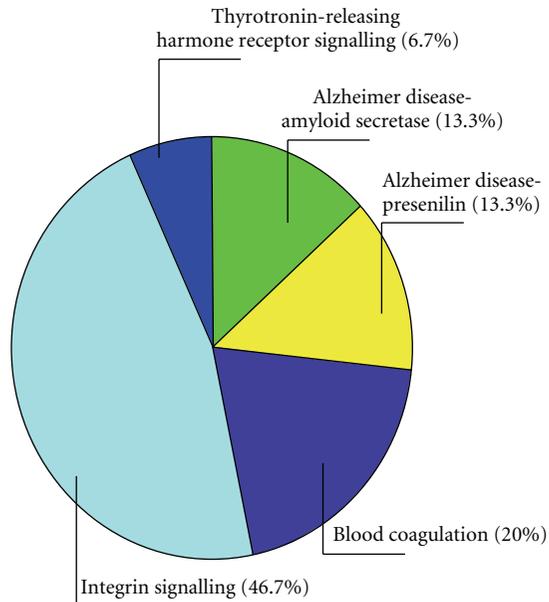
Figure 5: *PANTHER analysis for pathway*. In total, 28 proteins were identified as elevated and 22 proteins were under expressed. Shown above is the different signaling pathways hits by these protein.

4. Discussion

Quantitation of serum or plasma proteins via iTRAQ analysis has recently been suggested to be suitable for the detection of biomarkers, as the method is highly reproducible, with little run-to-run variation. This aspect, which is optimal to embark on such a "fishing-expedition", is actually quite surprising, granted the large number of individual steps in the work-flow. This conclusion was, however, derived at after a lengthy comparison involving three pools of case and control samples, as well as a number of individual samples, where a coefficient of variation of 11.7% was noted [21].

In a preliminary proof-of-principle experiment, we have now assessed whether this method could be suitable for the detection of biomarkers useful for DS screening. From our small-scale preliminary evaluation we identified over 200 proteins whose concentration was altered in the plasma of pregnancies with a DS fetus when compared to those with euploid fetuses.

It is of interest that βhCG is detected in the pool of proteins found to be elevated. This glycoprotein, which is produced early in pregnancy by the developing embryo and subsequently by the syncytiotrophoblast, forms the backbone of 1st and 2nd trimester screening strategies. In the 1st trimester, pregnancies at-risk of carrying a fetus affected by DS are identified on the basis of an almost 2 fold MoM (Multiples of the Median) elevation in βhCG. The presence of this important screening marker in our pool of elevated plasma proteins suggests that the strategy we have chosen for the identification of new biomarkers is functional and worthy of further pursuit.

In this context it is worth noting that this marker was not detected in two previous studies [9, 11] using proteomic

technologies for the identification of protein markers for DS in maternal plasma or serum. This could be due to limitations in detection sensitivity of the 2D gel approach these studies had used.

That our assay is indeed detecting proteins of placental origin is illustrated by presence of pregnancy zone protein which is a glycoprotein and a proteinase inhibitor which derive its importance in pregnancy by playing potential role in preventing the attack from maternal immune system on the developing fetus which can be seen as foreign allograft [22].This protein was amongst our list of up-regulated proteins, which is similar to alpha 2-macroglobulin, is quantitatively the most important pregnancy-associated plasma protein.

Amongst the pool of up-regulated proteins were serum amyloid P-component and amyloid beta A4, which is encouraging as members of this family have found to be elevated in previous studies on pregnancies with DS fetuses, which may be a reflection of an altered gene expression pattern associated with DS-related dementia. It is also important to note that the amyloid precursor protein (APP) gene is located in DS region of chromosome 21. Teller and colleagues have shown that a relatively minor from of amyloid beta protein was present in the brain of DS affected pregnancy as early as in late 2nd trimester [23].

Elevations in a number of inflammatory molecules, most of which are probably of maternal origin, may be a reflection of the elevated release of placental debris which has been suggested to occur in pregnancies with DS fetuses. Phosphatidylinositol-glycan is expressed in the placental tissue [24], also involved in many cellular process and plays important role in several signal transduction pathways.

Of the down-regulated proteins like actin, gelsolin, heparan sulfate proteoglycan [25], and fibrinogen alpha chain, it is noteworthy that all these protein are involved in cell adhesion, extracellular matrix, cell structure or muscle contraction, since DS fetuses are known to exhibit connective tissue abnormalities [26], the most pronounced of which is the increased neck fold (nuchal translucency). Since adult DS patients are prone to skeletal muscle deficiency, our finding concerning reduced titin plasma levels is intriguing, as this very large protein plays an important role in muscle contraction. In one of recent study Du and colleague [27] have shown trophoblast expression of titin in first trimester placentae.

Only a limited number of studies have attempted to use proteomic approaches for the discovery of new biomarkers for pregnancies at-risk of carrying a fetus with DS, two of these used 2-DE approaches [9, 11], while a further used a SELDI method [10]. Study of the former studies that by Nagalla and colleagues is the largest, having examined serum samples from 56 pregnant women. This study used samples collected in both the 1st and 2nd trimester of pregnancy, which were recruited as part of the NIH funded FASTER study, and largely made use of the fluorescent 2D-DIGE process. In their study, 18 proteins were found to be elevated in 1st trimester samples, which included members of the apoliprotein family, clusterin and proteins involved in skeletal development (tetranectin). The study by

Kolialexi and colleagues [11] used traditional 2DE stained with Coomassie blue on 20 maternal plasma samples (8 cases, 12 controls, 16–18 weeks of pregnancy), by which means 8 candidate proteins were detected. Elevations were noted for apoliprotein E and serum amyloid P-component. In contrast to the study by Nagalla et al. [9] a down regulation for clusterin was noted.

In the study by Busch and colleagues [10] using a SELDI approach, traces were noted which differed between DS cases and controls. However, no attempt was made to discern what proteins were responsible for these altered patterns, nor was any detailed description provided of how they could be reproduced.

Other than common elevation in serum amyloid and complement component families, little commonality exists between our study and these studies. This may be due to a number of factors including, limited study size, time of sampling collection, sample processing and storage, as well as use of very different technical approaches. In our follow-up studies we would like to validate these putative biomarkers using immunoblot and Enzyme linked immunosorbent assay (ELISA). More recently Selected Reaction Monitoring (SRM) [28] has evolved as a method of choice for validation of biomarkers using mass spectroscopy.

The increasing popularity of the iTRAQ approach due to its reproducibility and robustness, including studies for cancer or inflammatory autoimmune disorder specific biomarkers suggests that it will become the method of choice for future studies, until it is surpassed by a new technical development. As pregnancy represents a unique constellation, whereby a foreign being is supported and nourished by the host, it may serve as an ideal model for proteomic analyses, as any unique markers should ideally disappear post delivery. Furthermore, as very few specific biomarkers exist to assist with the screening of a number of pregnancy-related disorders, especially preeclampsia or preterm labour, it is likely that this will become the focus of considerable research attention in the near future.

5. Conclusion

In this report we conclude that isobaric labelling technique is a suitable approach for the quantitative detection of new screening biomarkers in the plasma of pregnancies with a DS fetus compared to those with euploid fetuses. In this preliminary proof-of-principle study, we were able to detect quantitatively under- or over-expressed proteins. In the future additional studies, using larger sample sizes will be required to identify a panel of biomarkers which can be used in screening for DS pregnancies.

Acknowledgment

References

[1] K. A. Eddleman, F. D. Malone, L. Sullivan, et al., "Pregnancy loss rates after midtrimester amniocentesis," *Obstetrics and Gynecology*, vol. 108, no. 5, pp. 1067–1072, 2006.

[2] E. Guetta, M. J. Simchen, K. Mammon-Daviko, et al., "Analysis of fetal blood cells in the maternal circulation: challenges, ongoing efforts, and potential solutions," *Stem Cells and Development*, vol. 13, no. 1, pp. 93–99, 2004.

[3] Y. M. D. Lo, "Recent advances in fetal nucleic acids in maternal plasma," *Journal of Histochemistry and Cytochemistry*, vol. 53, no. 3, pp. 293–296, 2005.

[4] R. D. Wilson, "Cell-free fetal DNA in the maternal circulation and its future uses in obstetrics," *Journal of Obstetrics and Gynaecology Canada*, vol. 27, no. 1, pp. 54–62, 2005.

[5] K. Spencer, "Aneuploidy screening in the first trimester," *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, vol. 145, no. 1, pp. 18–32, 2007.

[6] R. Ricco, A. M. Dalena, T. Valente, et al., "Quantitative study of placental villi in trisomy by analytical morphometry," *Analytical and Quantitative Cytology and Histology*, vol. 31, no. 1, pp. 41–48, 2009.

[7] S.-J. Park, W.-G. Yoon, J.-S. Song, et al., "Proteome analysis of human amnion and amniotic fluid by two-dimensional electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry," *Proteomics*, vol. 6, no. 1, pp. 349–363, 2006.

[8] G. T. Tsangaris, A. Kolialexi, P. M. Karamessinis, et al., "The normal human amniotic fluid supernatant proteome," *In Vivo*, vol. 20, no. 4, pp. 479–490, 2006.

[9] S. R. Nagalla, J. A. Canick, T. Jacob, et al., "Proteomic analysis of maternal serum in down syndrome: identification of novel protein biomarkers," *Journal of Proteome Research*, vol. 6, no. 4, pp. 1245–1257, 2007.

[10] A. Busch, S. Michel, C. Hoppe, D. Driesch, U. Claussen, and F. Von Eggeling, "Proteome analysis of maternal serum samples for trisomy 21 pregnancies using proteinchip arrays and bioinformatics," *Journal of Histochemistry and Cytochemistry*, vol. 53, no. 3, pp. 341–343, 2005.

[11] A. Kolialexi, A. Mavrou, G. Spyrou, and G. T. Tsangaris, "Mass spectrometry-based proteomics in reproductive medicine," *Mass Spectrometry Reviews*, vol. 27, no. 6, pp. 624–634, 2008.

[12] N. L. Anderson and N. G. Anderson, "The human plasma proteome: history, character, and diagnostic prospects," *Molecular & Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, 2002.

[13] P. L. Ross, Y. N. Huang, J. N. Marchese, et al., "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents," *Molecular & Cellular Proteomics*, vol. 3, no. 12, pp. 1154–1169, 2004.

[14] P. G. Righetti, E. Boschetti, L. Lomas, and A. Citterio, "Protein equalizer technology : the quest for a "democratic proteome"," *Proteomics*, vol. 6, no. 14, pp. 3980–3992, 2006.

[15] R. D. Unwin, A. Pierce, R. B. Watson, D. W. Sternberg, and A. D. Whetton, "Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells," *Molecular & Cellular Proteomics*, vol. 4, no. 7, pp. 924–935, 2005.

[16] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry," *Nature Methods*, vol. 4, no. 3, pp. 207–214, 2007.

[17] I. V. Shilov, S. L. Seymourt, A. A. Patel, et al., "The paragon algorithm, a next generation search engine that uses sequence temperature values sequence temperature values and feature

probabilities to identify peptides from tandem mass spectra," *Molecular & Cellular Proteomics*, vol. 6, no. 9, pp. 1638–1655, 2007.

[18] H. Mi, N. Guo, A. Kejariwal, and P. D. Thomas, "PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways," *Nucleic Acids Research*, vol. 35, database issue, pp. D247–D252, 2007.

[19] A. Glen, C. S. Gan, F. C. Hamdy, et al., "iTRAQ-facilitated proteomic analysis of human prostate cancer cells identifies proteins associated with progression," *Journal of Proteome Research*, vol. 7, no. 3, pp. 897–907, 2008.

[20] C. S. Gan, P. K. Chong, T. K. Pham, and P. C. Wright, "Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ)," *Journal of Proteome Research*, vol. 6, no. 2, pp. 821–827, 2007.

[21] X. Song, J. Bandow, and J. Sherman, "iTRAQ experimental design for plasma biomarker discovery," *Journal of Proteomic Research*, vol. 7, no. 7, pp. 2952–2958, 2008.

[22] E. L. Skornicka, N. Kiyatkina, M. C. Weber, M. L. Tykocinski, and P. H. Koo, "Pregnancy zone protein is a carrier and modulator of placental protein-14 in T-cell growth and cytokine production," *Cellular Immunology*, vol. 232, no. 1-2, pp. 144–156, 2004.

[23] J. K. Teller, C. Russo, L. M. DeBusk, et al., "Presence of soluble amyloid β-peptide precedes amyloid plaque formation in Down's syndrome," *Nature Medicine*, vol. 2, no. 1, pp. 93–95, 1996.

[24] A. M. Vinggaard, J. J. Provost, J. H. Exton, and H. S. Hansen, "Arf and RhoA regulate both the cytosolic and the membrane-bound phospholipase D from human placenta," *Cellular Signalling*, vol. 9, no. 2, pp. 189–196, 1997.

[25] R. Timpl and J. C. Brown, "Supramolecular assembly of basement membranes," *BioEssays*, vol. 18, no. 2, pp. 123–132, 1996.

[26] F. K. Wiseman, "Down syndrome—recent progress and future prospects," *Human Molecular Genetics*, vol. 18, no. R1, pp. R75–R83, 2009.

[27] M.-R. Du, W.-H. Zhou, F.-T. Yan, et al., "Cyclosporine A induces titin expression via MAPK/ERK signalling and improves proliferative and invasive potential of human trophoblast cells," *Human Reproduction*, vol. 22, no. 9, pp. 2528–2537, 2007.

[28] V. Lange, "Selected reaction monitoring for quantitative proteomics: a tutorial," *Molecular Systems Biology*, vol. 4, p. 222, 2008.

Research Article

Evaluation of Hepcidin Isoforms in Hemodialysis Patients by a Proteomic Approach Based on SELDI-TOF MS

Natascia Campostrini,[1] Annalisa Castagna,[1] Federica Zaninotto,[1] Valeria Bedogna,[2] Nicola Tessitore,[2] Albino Poli,[3] Nicola Martinelli,[1] Antonio Lupo,[2] Oliviero Olivieri,[1] and Domenico Girelli[1]

[1] *Department of Medicine, University of Verona, 37134 Verona, Italy*
[2] *Nephrology and Haemodialysis Unit, Division of Nephrology, University of Verona, 37134 Verona, Italy*
[3] *Department of Medicine and Public Health, University of Verona, 37134 Verona, Italy*

Correspondence should be addressed to Annalisa Castagna, annalisa.castagna@univr.it

The hepatic iron regulator hormone hepcidin consists, in its mature form, of 25 amino acids, but two other isoforms, hepcidin-20 and hepcidin-22, have been reported, whose biological meaning remains poorly understood. We evaluated hepcidin isoforms in sera from 57 control and 54 chronic haemodialysis patients using a quantitative proteomic approach based on SELDI-TOF-MS. Patients had elevated serum levels of both hepcidin-25 and hepcidin-20 as compared to controls (geometric means: 7.52 versus 4.69 nM, and 4.06 versus 1.76 nM, resp., $P < .05$ for both). The clearance effects of a single dialysis session by different dialysis techniques and membranes were also investigated, showing an average reduction by 51.3% ± 29.2% for hepcidin-25 and 34.2% ± 28.4% for hepcidin-20 but only minor differences among the different dialysis modalities. Measurement of hepcidin isoforms through MS-based techniques can be a useful tool for better understanding of their biological role in hemodialysis patients and other clinical conditions.

1. Introduction

Regulation of iron metabolism is crucial in different types of cells, especially in hepatocytes, enterocytes, and macrophages [1]. The major player of this regulation has been recognized in the small peptide hepcidin, that inhibits iron export from the cell membrane through internalization and degradation of ferroportin [2]. Hepcidin is a defensin-like peptide prevalently produced by the liver as an 84-amino acid precursor, that undergoes sequential proteolytic cleavage to form the 25-amino acid bioactive hormone (hepcidin-25). Additional N-terminal cleavage results in the production of two smaller isoforms, hepcidin-22 and hepcidin-20, whose physiological role is still unclear. Hepcidin-22 has been found only in urine, where it may merely represent a degradation product. On the other hand, hepcidin-20 is also present in blood and appears to possess greater antimicrobial activity than hepcidin-25 [3] but no capability to bind ferroportin [4]. Moreover, it has been recently indicated as a possible serum marker for the acute phase of myocardial infarction [5], but systematic investigations on this isoform in human diseases are lacking.

Until recently, studies on hepcidin have been limited by technical difficulties in establishing reliable assays in biological fluids [6]. Proteomic approaches including Surface-Enhanced Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (SELDI-TOF MS) are increasingly used for the evaluation of biomarkers of clinical interest [7, 8]. Indeed, we and others have contributed to the development and validation of an accurate assay of hepcidin in serum and urine by means of SELDI-TOF MS [9–12]. At variance with traditional immunochemical methods, the SELDI-TOF MS technique has the potential to evaluate not only the predominant 25-amino acid form, but also the shorter isoforms.

Hepcidin has been recently object of intense investigation as a potential biomarker of iron status in patients with chronic kidney diseases (CKD) [13]. Indeed, an adequate iron availability is critical in hemodialysis (HD) patients, since it appears to influence the response to recombinant erythropoietin (EPO), the mainstay of treatment of anemia in this condition. In HD patients, various factors may modulate serum hepcidin levels with opposing influence (reviewed in [14]). For example, hepcidin may increase because of reduced glomerular filtration, iron therapy, and inflammation, with interleukin-6 as a well-known stimulus for its production [13, 15]. On the other hand, hepcidin may be reduced by hypoxia, iron deficiency, and EPO therapy by itself [11]. Recent studies using different hepcidin assays have established that serum hepcidin-25 levels are generally increased in HD patients [16–19], but whether or not it can represent a useful marker to predict resistance to EPO remains controversial.

Moreover, little is known about the levels of the hepcidin isoforms in HD patients, as well as about the effect of a single dialysis session on the clearance of hepcidins from the blood.

The aims of this study were to evaluate serum levels of hepcidin-25 and hepcidin-20 by quantitative SELDI-TOF MS [9] in chronic HD patients, their relationship with markers of iron status and/or inflammation, and the effects of a single dialysis using different protocols on the clearance of hepcidin isoforms.

2. Materials and Methods

Fifty-four stable HD patients treated at the Hemodialysis Service of our hospital (Ospedale Policlinico, Verona, Italy) during the period June 2008–July 2009 were included in the study. Patients were on maintenance with intravenous erythropoiesis-stimulating agents (ESAs) epoetin or darbepoetin and chronically treated with intravenous iron, which was stopped 10 weeks before the study. All subjects were dialyzed against standard, not ultrapure, dialysate. Different protocols were used. Thirty-nine patients were on bicarbonate HD (BHD), 23 with low-flux synthetic (11 polysulfone, Fresenius, Germany and 12 polyamide, Gambro, Sweden), and 16 with high-flux membranes (6 cellulose triacetate, Nipro, Japan, 5 each polysulfone, Fresenius, Germany, and polymethylmethacrylate, Toray, Japan). Eleven patients were on acetate-free biofiltration (AFB), a low-volume hemodiafiltration technique based on buffer-free dialysate, a biocompatible high-flux AN69 membrane, and sterile hypertonic bicarbonate infusion in post-dilution mode (Hospal, France) [20]. Finally 4 patients were on HFR (double-chamber hemodiafiltration with reinfusion of regenerate ultrafiltrate), a technique that utilizes convection, diffusion, and adsorption, using a $0.7\,m^2$ high permeability polyphenylene membrane as a convective dialyzer, a $1.70\,m^2$ low-flux polyphenylene membrane as a diffusive dialyzer, and a regenerating adsorbent cartridge containing undissolvable macroporous-structured styrenic resin as adsorbent material (Bellco, Italy) [21]. Dialyzers were not reused. In all patients the length of the dialysis session was set at 240

TABLE 1: Clinical features and baseline laboratory data.

	HD patients	Controls
Number of patients	54	57
Gender (Male/Female)	29/25	38/21
Age (years)	67 ± 14	35 ± 15
HFE genotype		
wt/wt (n)	35	—
H63D/wt (n)	14	—
C282Y/wt (n)	2	—
Hemoglobin (g/dL)	11.4 ± 1.0	14.4 ± 1.2
Iron (μg/dL)	$45.2\ (43.1-60.0)$	$96.4\ (89.1-104.2)$
Transferrin Saturation (%)	20.5 ± 8.3	28.9 ± 9.0
Ferritin (ng/mL)	$174\ (156-258)$	$90\ (47-149)$
CRP (mg/dL)	$4.7\ (5.0-8.9)$	$0.8\ (0.4-1.4)$
Interleukin-6 (pg/mL)	$6.8 \pm 5.8^*$	—

*Reference range in normal subjects = 0.2–3.2 pg/mL.

minutes, and the dialysate flow rate at 500 mL/min, the dialysis blood pump flow rate at 300 mL/min, and in the patients on AFB, the 164 mM bicarbonate reinfusion fluid rate was at 2 L/h.

Blood samples for laboratory testing were obtained prior to the first-of-the-week hemodialysis sessions. Blood samples were collected for detection of C-reactive protein (CRP), ferritin, and interleukin-6 (IL-6) before the dialysis session. Serum ferritin was measured by routine laboratory methods, IL-6 by enzyme-linked immunoadsorbent assay (by Human IL-6 ELISA BMS213/2CE Bender MedSystems GmbH, Vienna, Austria), and CRP by commercially available automated PENIA assays (Dade Behring, Germany).

The effect of the dialysis session on serum hepcidin-25 and hepcidin-20 was evaluated in 39 patients (14 on low-flux BHD, 12 on high-flux BHD, 9 on AFB, and 4 on HFR) by measuring hepcidins also at the end of the dialysis sessions and by calculating the reduction ratio (RR) as follows: RR = (Cpre − Cpost/Cpre) × 100, where Cpre is the concentration at the start of dialysis and Cpost the concentration at the end of dialysis. Cpost was corrected for hemoconcentration due to ultrafiltration by multiplying the uncorrected Cpost for a correction factor computed as the ratio predialysis Hb/postdialysis Hb.

Fifty-seven controls were enrolled among healthy volunteers participating in a phase II trial at the Centre for Clinical Research of the Azienda Ospedaliera-Universitaria di Verona, as described previously in detail elsewhere [12]. Briefly, at enrollment they completed a questionnaire with specific items relevant to iron metabolism (i.e., any history of blood donations, previous pregnancy, menstrual losses, etc.) and were evaluated by laboratory studies including ferritin, CRP, liver function tests, and creatinine. To be considered as appropriate "normal controls" for the serum hepcidin assay, all these parameters were required to be normal.

The main clinical features of the subjects included in the study are shown in Table 1.

Each patient gave written informed consent. The study was conducted according to the principles contained in

the Declaration of Helsinki. The protocol was approved by the Institutional Review Board of the Azienda Ospedaliera-Universitaria di Verona.

2.1. Serum Hepcidin-25 and Hepcidin-20 Assay. We used a protocol based on PBSCIIc mass spectrometer and copper-loaded immobilized metal-affinity capture ProteinChip arrays (IMAC30-Cu^{2+}). Five μL of serum were applied to an IMAC30- Cu^{2+} surface that binds hepcidin based on its affinity to Cu^{2+} ions. The binding surface was equilibrated and washed with appropriate buffers according to the manufacturs's instructions (Bio-rad, Hercules, CA). Subsequent work-up, SELDI-TOF MS instrumental settings, read-out, and data analysis are described elsewhere [12], with the addition that protein chip handling was performed in a nitrogen atmosphere to prevent methionine oxidation [9].

Briefly, we used synthetic hepcidin-25 (Peptides International, Louisville, KY) for external calibration and a synthetic hepcidin analogue (Hepdicin-24, Peptides International, Louisville, KY) as an internal standard [9]. Spectra were collected in duplicate for each sample, with or without the internal standard spiked in at a concentration of 10 nM. Concentrations of both serum hepcidin-25 and hepcidin-20 were expressed as nM and were the results of the following equations.

(1) Hepcidin 25 concentration: (sample 2789 m/z peak intensity) $\times$ 10 nM/(hepc24 spiked sample 2673 m/z peak intensity—nonspiked sample 2673 m/z peak intensity).

(2) Hepcidin 20 concentration: (sample 2192 m/z peak intensity) $\times$ 10 nM/(hepc24 spiked sample 2673 m/z peak intensity—nonspiked sample 2673 m/z peak intensity).

Standard curves of the internal standard were constructed by serial dilutions of hepcidin-24 (0–20 nM) in tubes with blank serum to an end volume of 500 μL. These were immediately applied to IMAC-Cu^{2+} Chips, and processed according to protocol and measured by MS. Linear standard curves were obtained for hepcidin-24 blank serum ($y = 4.90\times + 3.97$; $R^2 = 0.994$). Based on the measured background noise in each MS spectrum for serum samples, the lower limit of detection (LLOD) ranged from 0.55 to 1.55 nM. Since this method is based on the level of hepcidin-25 peak intensity relative to that of hepcidin-24, we determined hepcidin-24/hepcidin-25 intensity ratios in blank serum samples spiked with both synthetic compounds in duplicate of 8 different concentrations. As compared to our original description of the method [9], recent technical improvements allowed us to increase the mean peak ratios hepcidin-24/hepcidin-25 from 0.71 to 0.93.

Protocols for identification of peaks by immunological and mass spectrometry approaches have been described in detail elsewhere [22]. Hepcidin-25 and hepcidin-20 concentrations of 0.55 nM were arbitrarily assigned to samples with undetectable serum levels of hepcidin isoforms. The intra- and inter-assay coefficient of variations of this method

TABLE 2: Correlations between serum hepcidin-25 levels, inflammation indices, and serum hepcidin-20 in HD patients.

	Correlation coefficient	P
Ferritin	0.478	.003
IL-6	0.096	.572
CRP	0.016	.910
Hepcidin-20	0.575	<.001

TABLE 3: Correlations between serum hepcidin-20 levels, inflammation indices, and serum hepcidin-25 in HD patients.

	Correlation coefficient	P
Ferritin	0.240	.153
IL-6	0.163	.335
CRP	0.093	.505
Hepcidin-25	0.575	<.001

ranged from 6.1 to 7.3 percent, and from 5.7 to 11.7 (mean 7.7) percent, respectively, [9].

2.2. Statistical Analysis. All calculations were performed using the SPSS 16.0 software (SPSS Inc., Chicago, IL, USA). As some of the continuous variables of interest, including serum hepcidin and ferritin, showed a non-Gaussian distribution, their values were log-transformed and expressed as geometric means with 95% Confidence Intervals (CIs). For other variables, results were expressed also as means $\pm$ SD unless otherwise indicated. Correlations between quantitative variables were calculated by Pearson r test. Results were considered significant when P was <.05 (two tailed).

3. Results

Figure 1 shows the typical spectra obtained for the region where hepcidin peptides are detectable. The corresponding peaks of both hepcidin-20 and hepcidin-25 are visible, along with the peak of the added standard. Concentrations of hepcidin-20 and hepcidin-25 were calculated for each patient (before and after dialysis, when available) and control. Both hepcidin-25 and hepcidin-20 were significantly higher in HD patients as compared to controls (Figures 2 and 3): geometric means with 95% confidence intervals were 7.52 nM (5.39–10.48) in HD patients versus 4.69 nM (3.79–5.81) in controls for hepcidin-25, and 4.06 nM (3.45–4.77) in HD patients versus 1.76 nM (1.32–2.35) in controls for hepcidin-20, P < .05 for both.

Of note, hepcidin-20 was undetectable in 22 out of 57 healthy subjects, while it was always detectable in HD patients.

Tables 2 and 3 show the correlation coefficients of hepcidin-25 and hepcidin-20 with serum ferritin and markers of inflammation. Hepcidin-25 was positively correlated with serum ferritin, but not with inflammatory markers. On the other hand, hepcidin-20 was not significantly correlated
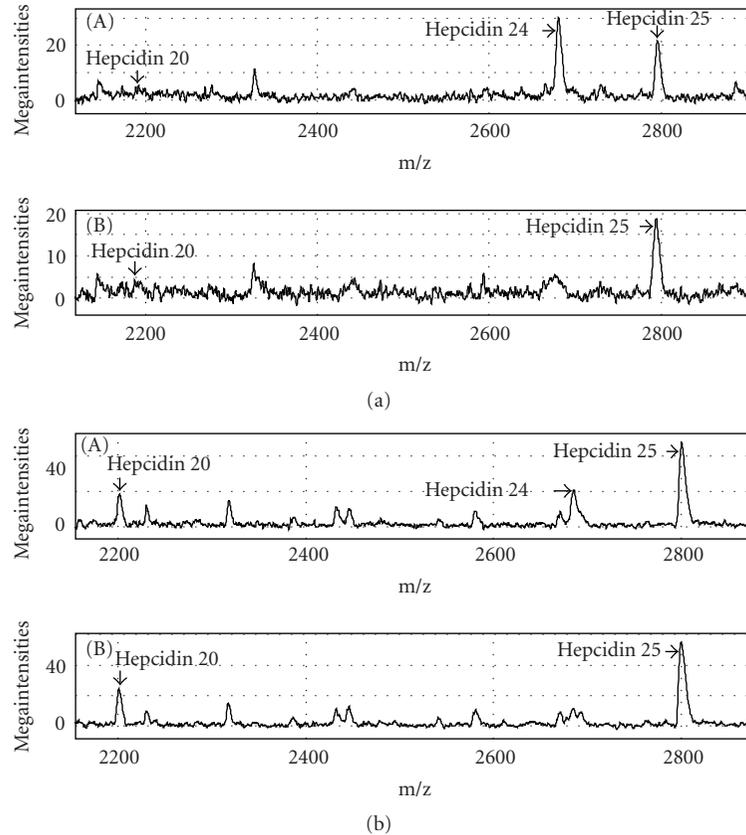
FIGURE 1: SELDI-TOF MS profile of hepcidin-24-spiked serum samples of hemodialysis patient (a) and control (b). The hepcidin isoforms hepcidin-20, hepcidin-24 (synthetic analogue, panel (B)), and hepcidin-25 are indicated by arrows.
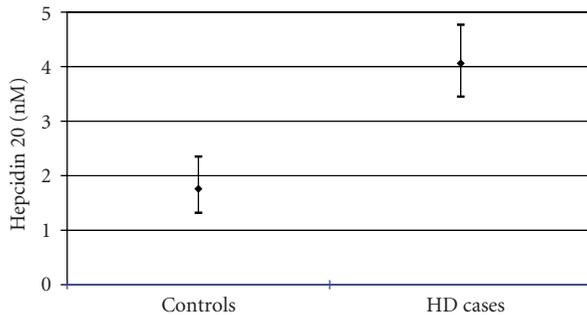

FIGURE 2: Hepcidin-20 in controls and HD patients. Data presented as geometric mean with 95% confidence interval.
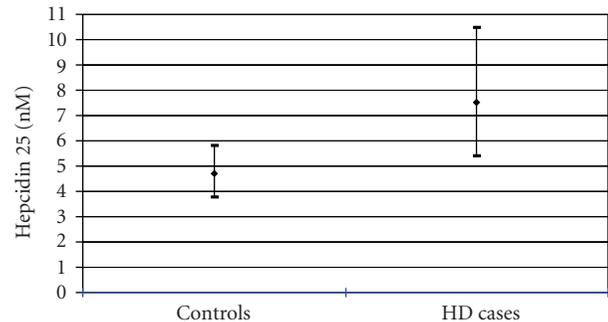

FIGURE 3: Hepcidin-25 in controls and HD patients. Data presented as geometric mean with 95% confidence interval.

with either ferritin or inflammatory markers, but was correlated positively with serum hepcidin-25.

The effects of a single dialysis session by the different dialysis techniques and membranes on serum hepcidin isoforms are shown in Figure 4. On average, we observed a reduction of 51.3%± 29.2% for hepcidin-25, and of 34.2%± 28.4% for hepcidin-20. The hepcidin-25 reduction ratio was similar by low- and high-flux BHD; removal by BHD was also similar to AFB and HFR. On the other hand, hepcidin-25 removal by the AFB was significantly lower than by HFR ($P = .033$) (Figure 4(a)). With respect to hepcidin-20 removal, no

significant difference was noted using the different dialysis techniques (Figure 4(b)). Of note, there was no difference in predialysis serum levels of both isoforms between the different dialysis modalities (Table 4).

4. Discussion

In this study, SELDI-TOF MS was confirmed as a valuable approach for the quantification of hepcidin levels in serum, not only with respect to the main isoform, hepcidin-25, but

TABLE 4: Predialysis serum hepcidin-25 and hepcidin-20 levels and the different HD techniques.

	Low-flux BHD ($n = 23$)	High-flux BHD ($n = 16$)	AFB ($n = 11$)	HFR ($n = 4$)
Hepcidin-25 (nM)	8.0 (7.7–16.2)	7.4 (6.3–16.0)	7.4 (6.0–22.9)	5.7 (0–19.5)
Hepcidin-20 (nM)	4.4 (3.7–6.8)	3.8 (3.1–5.9)	4.2 (2.9–7.4)	2.7 (1.3–4.3)

Data are presented as geometric mean (95% Confidence Interval).
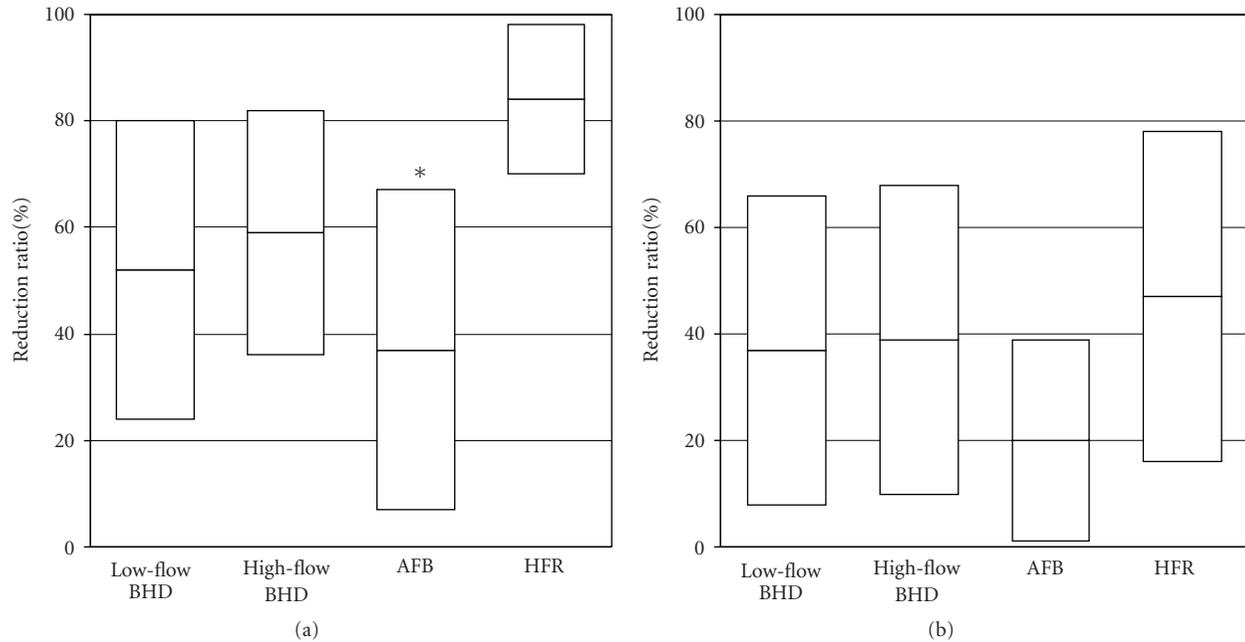
FIGURE 4: Reduction Ratio (%) of serum hepcidin-25 (a) and hepcidin-20 (b) by the different hemodialysis techniques. (*Statistically significant $P < .05$).

also to hepcidin-20. According to recent data obtained with various methodological approaches, including competitive ELISA [19], radioimmunoassay [17], micro-HPLC-tandem-MS [18], as well as this method [18], hepcidin-25 was proven to be increased in HD patients as compared to control subjects. Two studies [17, 19] also observed a gradual increase of hepcidin-25 across the spectrum of predialysis CKD, suggesting that diminished renal clearance is likely the main cause of elevated serum hepcidin-25 in these pathological conditions. The positive correlation with serum ferritin, confirmed in this study, indicates that the physiological regulation of hepcidin-25 by iron stores is maintained in CKD, though at an upper level as compared to controls subjects. Besides these confirmatory results, this study focused on two relatively unexplored aspects of hepcidin pathophysiology in HD, for example, the evaluation of serum levels of hepcidin-20 and the effects of different dialysis procedures on blood levels of hepcidin isoforms.

4.1. Baseline Serum Hepcidin-20 Levels in HD Patients. We evaluated for the first time serum hepcidin-20 levels in both normal subjects and HD patients. On average, hepcidin-20 was significantly higher in HD patients as compared to controls. More precisely, hepcidin-20 was undetectable in a substantial fraction (near 38%) of healthy subjects but

was always detectable in HD patients. While hepcidin-20 was significantly and positively correlated with hepcidin-25, it was not correlated with either serum ferritin or markers of inflammation, such as CRP and IL-6. No obvious pathophysiological role of hepcidin-20 can be inferred by these results. The lack of correlation with iron stores and/or inflammation, for example, the classical stimuli for the iron bioactive hepcidin-25 isoform, argues against a role of hepcidin-20 in these conditions. For example, if we had noted a positive relationship with CRP and/or IL-6, one could speculate about a selective, inflammatory-driven, posttranslational modification of hepcidin to increase the relative fraction of the isoform with the highest reminiscent antimicrobial activity. Rather, our results may indicate two contributing mechanisms as responsible of the increased hepcidin-20 levels in HD: an altered proteolytic processing and/or a reduced renal excretion of a catabolic product normally produced in small amount and nearly completely eliminated. In line with this interpretation is the positive correlation we observed between hepcidin-20 and hepcidin-25, suggesting a "tracking effect" of the main isoform. While the pathophysiological role of hepcidin-20 in HD, if any exists, remains to be elucidated, the significant increase we noted is not without practical relevance. Indeed, as pointed out in a recent review [23], while different methods uniformly found

high hepcidin levels in HD, absolute values varied as much as 10-fold depending on the hepcidin assay used. In particular, the immunological methods generally reported the highest levels of "hepcidin-25" as compared to MS-based methods, including the present SELDI-TOF MS. This may be due to, at least in part, cross-reaction of antibodies with hepcidin-20. Since in terms of iron homeostasis hepcidin-25 is the sole bioactive form, MS-based methods may be more accurate for clinical investigations, notwithstanding being more complex than traditional immunological methods.

4.2. Effects of Different Dialysis Procedures on Blood Levels of Hepcidin Isoforms. The dialysis session appeared to significantly reduce average hepcidin-25 levels, as also reported by others using MS-based methods [18], but not immunological methods [17]. We also observed that the extent of the reduction by dialysis showed a high interindividual variability (see below) [24]. Of note, the lack of hepcidin-25 reduction after dialysis reported with [17] an immunological method was based on results in only 6 HD patients, suggesting the large individual variability as a possible confounder, beyond the effect of the different methodology employed.

All HD techniques evaluated in our study showed a comparable removal of hepcidin-25 by dialysis, suggesting that the removal of this molecule is largely due to diffusion. The most striking result of our study is indeed the large individual variability of the reduction ratio in patients treated by the same dialysis modality (ranging from 0 to 95%), suggesting that postdialysis levels may be influenced by factors other than removal by dialysis. One such factor may be an increased production of hepcidin-25 during the dialysis procedure, which is known to activate an inflammatory response due to the interaction of blood with the dialysis membranes and/or with a less-than-sterile dialysis fluid [25]. This possibility is supported by the finding that the greatest removal was observed by HFR, a dialysis technique associated with more efficient removal of cytokines and reduced levels of markers of inflammation (such as CRP and IL-6) when compared to standard hemodialysis [21]. Alternatively, the improved reduction ratio by HFR may be due to adsorption, which may provide an important additional modality for hepcidin-25 removal by dialysis. In addition, mean predialysis serum hepcidin-25 levels were lower in patients on long-term HFR (although this difference was not statistically significant), again suggesting that dialysis modalities with lower activation of the inflammatory response may contribute to reduce the hepcidin-25 burden.

In our study, removal of hepcidin-25 by BHD with low- and high-flux membranes was similar, at variance with what reported by Weiss and coworkers [18], who showed that postdialysis hepcidin-25 levels were lower with high-flux than low-flux BHD, suggesting an improved removal by high-flux membranes. These findings, however, may not be necessarily conflicting, since the effects of dialysis were evaluated differently in the two studies. Weiss and coworkers reported average postdialysis serum levels, while we computed the reduction ratio by dialysis. In addition,

the differences between the two studies may be explained by the large intraindividual variability of the effect of dialysis reported by both and their small sample size.

Although the removal of hepcidin-20 by the different dialysis modalities was slightly lower than hepcidin-25, this difference was not statistically significant, suggesting that the different hemodialysis modalities remove the two isoforms by the same extent and through common mechanisms.

We are aware that the results of our study on the effect of the different dialysis techniques on circulating levels of hepcidin-25 and hepcidin-20 are preliminary and should be considered with caution. Clearly, additional studies with larger sample size and more adequate design are needed to fully understand the role of the dialysis procedure on serum levels of the two hepcidin isoforms.

5. Conclusions

Proteomics can be very useful in molecular medicine nowadays, as precious tools for biomarker discovery and quantitation. SELDI-TOF MS proteomic analysis enabled us to evaluate hepcidin isoforms in sera from HD patients before and after dialysis, giving new insights into the complex pathophysiology of hepcidin in chronic kidney disease.

Abbreviations

SELDI-TOF MS:	Surface-Enhanced Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry
CKD:	Chronic kidney disease
HD:	Hemodialysis
EPO:	Erythropoietin
ESA:	Erythropoiesis-stimulating agents
BHD:	Bicarbonate HD
AFB:	Acetate-free biofiltration
HFR:	Double chamber hemodiafiltration with reinfusion of regenerate ultrafiltrate
CRP:	C-reactive protein
IL-6:	Interleukin-6
CIs:	Confidence intervals.

Acknowledgments

References

[1] T. Ganz, "Hepcidin, a key regulator of iron metabolism and mediator of anemia of inflammation," *Blood*, vol. 102, no. 3, pp. 783–788, 2003.

[2] E. Nemeth, M. S. Tuttle, J. Powelson, et al., "Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization," *Science*, vol. 306, no. 5704, pp. 2090–2093, 2004.

[3] C. H. Park, E. V. Valore, A. J. Waring, and T. Ganz, "Hepcidin, a urinary antimicrobial peptide synthesized in the liver,"

Journal of Biological Chemistry, vol. 276, no. 11, pp. 7806–7810, 2001.

[4] I. De Domenico, E. Nemeth, J. M. Nelson, et al., "The hepcidin-binding site on ferroportin is evolutionarily conserved," *Cell Metabolism*, vol. 8, no. 2, pp. 146–156, 2008.

[5] H. Suzuki, K. Toba, K. Kato, et al., "Serum hepcidin-20 is elevated during the acute phase of myocardial infarction," *Tohoku Journal of Experimental Medicine*, vol. 218, no. 2, pp. 93–98, 2009.

[6] A. Piperno, R. Mariani, P. Trombini, and D. Girelli, "Hepcidin modulation in human diseases: from research to clinic," *World Journal of Gastroenterology*, vol. 15, no. 5, pp. 538–551, 2009.

[7] M. Kiehntopf, R. Siegmund, and T. Deufel, "Use of SELDI-TOF mass spectrometry for identification of new biomarkers: potential and limitations," *Clinical Chemistry and Laboratory Medicine*, vol. 45, no. 11, pp. 1435–1449, 2007.

[8] J. A. Bons, W. K. Wodzig, and M. P. van Dieijen-Visser, "Protein profiling as a diagnostic tool in clinical chemistry: a review," *Clinical Chemistry and Laboratory Medicine*, vol. 43, no. 12, pp. 1281–1290, 2005.

[9] D. W. Swinkels, D. Girelli, C. Laarakkers, et al., "Advances in quantitative hepcidin measurements by time-of-flight mass spectrometry," *PLoS ONE*, vol. 3, no. 7, article e2706, 2008.

[10] D. Girelli, I. De Domenico, C. Bozzini, et al., "Clinical, pathological, and molecular correlates in ferroportin disease: a study of two novel mutations," *Journal of Hepatology*, vol. 49, no. 4, pp. 664–671, 2008.

[11] P. Robach, S. Recalcati, D. Girelli, et al., "Alterations of systemic and muscle iron metabolism in human subjects treated with low-dose recombinant erythropoietin," *Blood*, vol. 113, no. 26, pp. 6707–6715, 2009.

[12] C. Bozzini, N. Campostrini, P. Trombini, et al., "Measurement of urinary hepcidin levels by SELDI-TOF-MS in HFE-hemochromatosis," *Blood Cells, Molecules, and Diseases*, vol. 40, no. 3, pp. 347–352, 2008.

[13] J. Malyszko and M. Mysliwiec, "Hepcidin in anemia and inflammation in chronic kidney disease," *Kidney and Blood Pressure Research*, vol. 30, no. 1, pp. 15–30, 2007.

[14] Y. Hamada and M. Fukagawa, "Is hepcidin the star player in iron metabolism in chronic kidney disease," *Kidney International*, vol. 75, no. 9, pp. 873–874, 2009.

[15] E. Nemeth, S. Rivera, V. Gabayan, et al., "IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin," *Journal of Clinical Investigation*, vol. 113, no. 9, pp. 1271–1276, 2004.

[16] L. Valenti, D. Girelli, G. F. Valenti, et al., "HFE mutations modulate the effect of iron on serum hepcidin-25 in chronic hemodialysis patients," *Clinical Journal of the American Society of Nephrology*, vol. 4, no. 8, pp. 1331–1337, 2009.

[17] D. R. Ashby, D. P. Gale, M. Busbridge, et al., "Plasma hepcidin levels are elevated but responsive to erythropoietin therapy in renal disease," *Kidney International*, vol. 75, no. 9, pp. 976–981, 2009.

[18] G. Weiss, I. Theurl, S. Eder, et al., "Serum hepcidin concentration in chronic haemodialysis patients: associations and effects of dialysis, iron and erythropoietin therapy," *European Journal of Clinical Investigation*, vol. 39, no. 10, pp. 883–890, 2009.

[19] J. Zaritsky, B. Young, H.-J. Wang, et al., "Hepcidin—a potential novel biomarker for iron status in chronic kidney disease," *Clinical Journal of the American Society of Nephrology*, vol. 4, no. 6, pp. 1051–1056, 2009.

[20] P. Zucchelli, A. Santoro, and G. Raggiotto, "Biofiltration in uremia: preliminary observations," *Blood Purification*, vol. 2, no. 4, pp. 187–195, 1984.

[21] M. L. Wratten and P. M. Ghezzi, "Hemodiafiltration with endogenous reinfusion," *Contributions to Nephrology*, vol. 158, pp. 94–102, 2007.

[22] E. H. J. M. Kemna, H. Tjalsma, V. N. Podust, and D. W. Swinkels, "Mass spectrometry-based hepcidin measurements in serum and urine: analytical aspects and clinical implications," *Clinical Chemistry*, vol. 53, no. 4, pp. 620–628, 2007.

[23] B. Young and J. Zaritsky, "Hepcidin for clinicians," *Clinical Journal of the American Society of Nephrology*, vol. 4, no. 8, pp. 1384–1387, 2009.

[24] N. Tomosugi, H. Kawabata, R. Wakatabe, et al., "Detection of serum hepcidin in renal failure and inflammation by using ProteinChip System," *Blood*, vol. 108, no. 4, pp. 1381–1387, 2006.

[25] G. A. Kaysen, "The microinflammatory state in uremia: causes and potential consequences," *Journal of the American Society of Nephrology*, vol. 12, no. 7, pp. 1549–1557, 2001.