CANCER GENOMICS & PROTEOMICS *3*: 11-28 (2006)

Review

Proteomics and Bioinformatics in Biomedical Research

THOMAS KISLINGER[1] and IGOR JURISICA[2,3]

[1]*Program in Proteomics and Bioinformatics, Banting and Best Institute of Medical Research and*
[2]*Departments of Computer Science and Medical Biophysics, University of Toronto, ON;*
[3]*Ontario Cancer Institute, Toronto Medical Discovery Tower, Toronto, ON, Canada*

Abstract. *Proteomics, the science of globally detecting proteins in cells, tissues or organisms under defined conditions, has greatly benefited from recent developments in mass spectrometry (MS). It is now possible to detect hundreds to thousands of proteins with high confidence in a single experiment. In this review, the basic MS technologies, currently used by laboratories around the world to identify proteins in complex biological samples, are summarized. Further, a short overview of useful separation strategies to minimize the initial complexity of biological samples, and the multitude of bioinformatics tools essential to manage large-scale proteomics data to obtain meaningful biological insight, is provided. Finally, recent advances in three main areas of medical proteomics are summarized: proteomics in cancer research, proteomics of the heart and proteomics in diabetes research.*

Proteome research has become a main stream research discipline over the last couple of years (1, 2). This is mainly due to the recent completion of the genomic sequences of several important organisms, providing proteome researchers with extensive protein sequence databases necessary for large-scale protein identifications (3-7). Likewise, significant advances in mass spectrometry over the past decade now make it possible to systematically identify thousands of proteins in even the most complex samples (Figure 1).

In this review, a systematic overview of commonly used mass spectrometers, sample separation methodologies and data mining strategies, currently used in modern proteomics laboratories is provided. Then, attention is focused on mass spectrometry-based proteomics in medical research, covering areas such as cancer proteomics, heart disease and diabetes.

Abbreviations: 1-DE, one-dimensional electrophoresis; 2-DE, two-dimensional electrophoresis; AGE, advanced glycation end-product; CE, capillary electrophoresis; CID, collision-induced dissociation; ESI, electrospray ionization; FT-ICR, Fourier transform ion cyclotron; GO, Gene Ontology; LC, liquid chromatography; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry: MS/MS, tandem mass spectrometry; MudPIT, multidimensional protein identification technology; PSLT, protein subcellular localization tool; RP, reverse phase; SELDI, surface-enhanced laser desorption/ionization.

Correspondence to: Thomas Kislinger, Ph.D., Program in Proteomics and Bioinformatics, Banting and Best Institute of Medical Research, University of Toronto, Ontario M5R 2W7, Canada. Tel: 416-946-5518, e-mail: thomas.kislinger@utoronto.ca

Key Words: Proteomics, mass spectrometry, bioinformatics, cancer biology, heart disease, diabetes, review.

Mass Spectrometry in Proteome Research

Mass spectrometry (MS) has long been used in analytical chemistry for the structural analysis of low molecular weight, volatile molecules. Although, not until the introduction of mild ionization technologies, such as electrospray ionization (ESI) (8) and matrix-assisted laser desorption/ionization (MALDI) (9) in the late 80's did MS enter into proteome research. These ionization technologies allow for a relatively mild ionization of large, intact biomolecules such as proteins, peptides, lipids and complex carbohydrates. The invention of these methodologies, greatly beneficial to modern proteomics, has been recognized with the 2002 Nobel Prize in Chemistry.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS)

In MALDI-TOF-MS protein or peptide samples are mixed in acidic solution with an excess of matrix (1:1000) and crystallized on a steel target. The matrix solution consists, in general, of an energy absorbing aromatic molecule, such as α-cyano-4-hydroxybenzoic acid. Ions are generated by pulsing the sample/matrix co-crystal with a nitrogen laser,

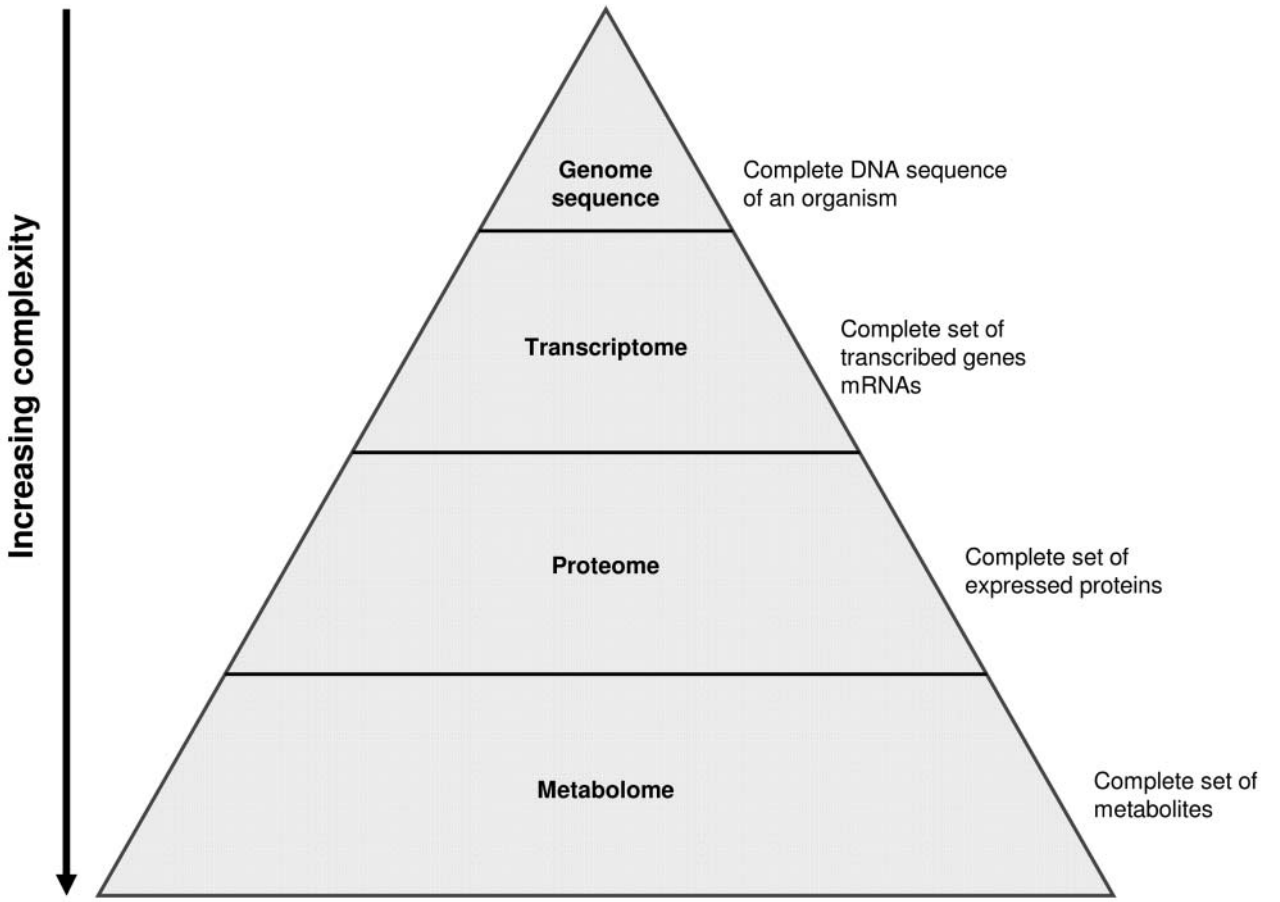

Figure 1. *Increasing layers of complexity in large-scale "omics" research. An increasing complexity from the DNA sequence to the number of expressed genes (mRNA), the number of expressed proteins, to the complete set of metabolites present is observed in large-scale biology.*

resulting in ionization of the sample and matrix molecules. A high potential electric field is used to accelerate the generated ions into the mass spectrometer. Traditionally, MALDI-TOF-MS-based proteomics experiments have been coupled to electrophoretic separation of proteins (see below) followed by in-gel digest of selected protein bands. The generated peptides are then analyzed by MALDI-TOF-MS to generate a "peptide fingerprint". In the case of highly purified proteins, especially for organisms such as yeast, this methodology has been used to great effect (10). Although, in the case of more complex samples or higher organisms, such as mouse or human, unambiguous protein identification is rarely the case. Nevertheless, the recent commercial introduction of MALDI-TOF/TOF-MS, capable of recording tandem mass spectra, has overcome some of these earlier limitations, especially if coupled to chromatographic separations (LC MALDI-TOF/TOF-MS). In clinical diagnostics, SELDI-TOF-MS (Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry) has recently made its entry. SELDI-TOF-MS is basically a variation of a MALDI-TOF-MS with the difference that the target plate is coated with chromatography resins (*e.g.*, ion-exchange, hydrophobic, metal ion binding, *etc.*). Although this technology offers high expectations in clinical diagnostics and biomarker discovery, unambiguous protein identification based on molecular weight alone cannot be achieved. For more detailed information on MALDI-TOF-MS, readers are encouraged to consult recently published reviews or specialized book chapters (11-17).

Electrospray ionization mass spectrometry (ESI-MS)

In ESI-MS, ions are formed as a fine spray of charged droplets at the tip of a fine needle in the presence of a strong electric field (2-5 kV). Quick evaporation of the solvent molecules results in the formation of gas phase ions, which are transferred on-line into the MS. As ions are formed directly from solution (in general acidic), ESI could be easily coupled to high-resolution separation technologies such as liquid chromatography (LC) or capillary electrophoresis (CE), making it an ideal methodology for the separation of

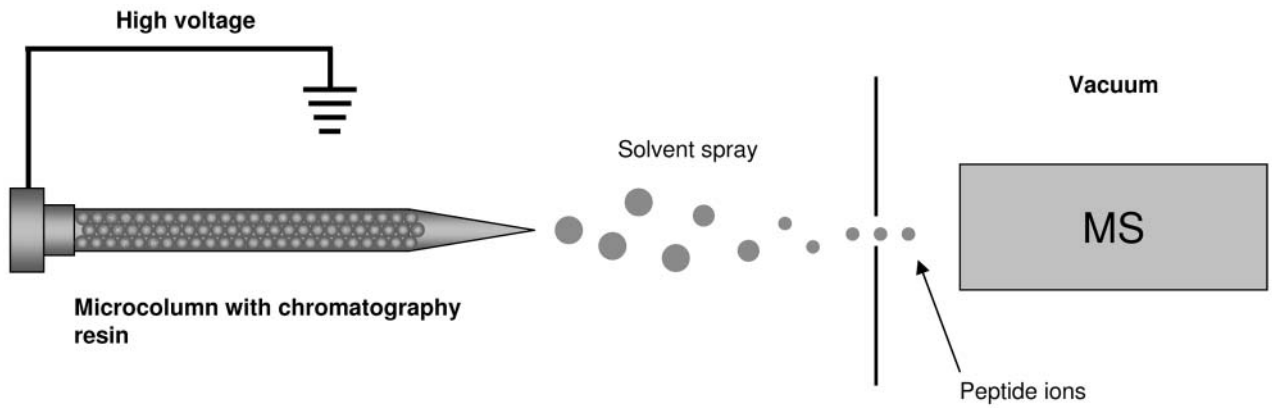

Figure 2. *Basic principle of LC-MS-based proteomics. Microcapillary columns are packed with a chromatography resin. Peptide samples are loaded onto the column and separated by solvent gradients. A high voltage is applied directly to the mobile phase and peptides are ionized and transferred directly into the mass spectrometer as they elute from the tip of the column.*

complex peptide samples. Particularly, the development of micro- and nano-electrospray ionization dramatically increased the sensitivity of peptide identifications in LC-MS-based proteomics. The basic advantage of nanospray ESI-MS is the reduction of the solvent flow rates to nanoliter per minute scales in narrow (~50-150 μm inner diameter) columns (see below for LC-MS) (Figure 2). The generated peptide ions are transferred into the MS and the mass over charge ratios (m/z) are systematically recorded up to a mass limit of approximately 2000 mass units. Individual ions are selectively isolated and fragmented (in general through collision with an inert gas molecule such as helium) in a process called collision-induced dissociation (CID). The result is a tandem mass spectrum (MS/MS-spectrum) which is distinct to the amino acid sequence of the fragmented peptide ion. The most commonly used mass spectrometers for ESI-MS, are the quadrupole-TOF and ion-trap instruments, which allow for the fast and reliable detection of several thousand MS/MS spectra in a single experiment. More recently, new mass analyzers, such as the Fourier transform ion cyclotron mass spectrometers (FT-ICR-MS) or the Orbitrap mass spectrometers, have been commercially introduced. These mass spectrometers provide a significantly improved mass accuracy (~1 ppm) and resolution, and should be particularly beneficial for specific applications in proteome research (*e.g.* mapping of posttranslational protein modifications). A considerable drawback of these mass spectrometers is their high initial and maintenance costs, clearly limiting these machines to more established or specialized investigators. Readers of this review are encouraged to consult the numerous excellent reviews and book chapters on the basics and principles of ESI-MS, tandem mass spectrometry and the variety of different mass spectrometers (ion-traps, FT-ICR-MS *etc.*) (1, 2, 14, 16).

Liquid chromatography mass spectrometry (LC-MS)

As mentioned above, in ESI-MS peptide ions are formed directly from solution, conveniently allowing for direct coupling to high-resolution separation technologies. In modern proteomics laboratories, narrow glass capillaries (inner diameters 50-150 μm) are used as chromatography columns, which are pulled to fine tips (diameter of ~5 μm), with laser pullers. The columns are custom-packed with a wide array of chromatography resins (*e.g.*, reverse phase RP-18 or ion exchange resins) using pressure vessels. Columns are placed in line with an HPLC pump and the peptide mixtures are separated and directly eluted into the MS with a solvent follow rate of several hundred nanoliters per minute (14). The mass spectrometer records the m/z of the peptides as they elute from the column over time. Importantly, the MS cycles between a MS-scan, simply recording m/z values of eluting peptides, and consecutive MS/MS-scans, in a data-dependent manner. Briefly, the most abundant peaks in each MS-scan are automatically isolated and fragmented by CID, resulting in sequence-dependent tandem mass spectra. Modern mass spectrometers, such as the popular linear ion-trap, are capable of recording several thousand MS/MS spectra in a single one-hour experiment. An inherent problem of LC-MS-based proteomics, especially in higher eukaryotic organisms, is the overwhelming complexity of the proteome, making it impossible to identify every protein expressed in a given sample (18, 19). To tackle this problem, several strategies have been applied in the literature to increase the number of identified proteins.

Multidimensional protein identification technology – MudPIT

A very powerful and convenient method to improve the detection coverage was published by Yates and colleagues in 2001, termed multidimensional protein identifications

technology (MudPIT) (20, 21). Basically, microcapillary columns are packed with two orthogonal chromatography resins, a reverse phase 18 resin (RP-18) and a strong cation exchange resin. Complex protein samples are then enzymatically digested with endoproteinase Lys-C and trypsin to very complex peptide mixtures. A multi-step sequence is used to separate and elute peptides directly into the MS. Briefly, a sequence consists of 5 or more independent chromatography steps which start with a "salt bump" of ammonium acetate to move a subset of the peptides from the ion exchange resin onto the RP-18 resin, followed by separation of these peptides by a conventional water-acetonitrile gradient. In the next step, the concentration of the "salt bump" is increased to move another set of peptides from the ion exchange resin onto the RP-18 resin (22, 23). The MudPIT technology has been used to great success in the literature and routinely identifies up to a thousand proteins in a single run (20, 21, 24-27). However, even high-resolution separation methods like MudPIT are not able to detect every eluting peptide in a complex mammalian protein extract (random sampling effect) (19). This fact can be partially overcome by analyzing a sample multiple times until a certain saturation of detection is achieved (18, 19, 28).

Alternative separation strategies

A multitude of protein and peptide separation technologies exists in analytical biochemistry, and several have been successfully applied to proteomics. Basically, all separation strategies aim at reducing the initial sample complexity to increase the number of detected proteins (detection coverage) or the biological information obtained from the experiment.

a) Organelle fractionation. Eukaryotic cells are segregated and composed into multiple organelles which perform dedicated functions within the cell. Isolation of organelles from eukaryotic cells or tissues by biochemical methods, such as ultracentrifugation in density gradients, has several advantages (29). First, the sample complexity of the protein extract is minimized as compared to the whole tissue extract. Second, potential biological information for unannotated proteins is obtained based on the subcellular localization. For example, an unannotated protein specifically detected in the nucleus most probably has a function related to this organelle. Several high impact papers have been published in recent years combining this strategy with mass spectrometry-based proteomics (25, 27, 28, 30-32).

b) Chromatographic fractionations. Besides the described on-line fractionations (*e.g.*, MudPIT), proteins or peptides can also be fractionated off-line. For example, intact proteins from a cell or tissue extract can be fractionated by ion-exchange chromatography, gel filtration, or other established chromatographic separations (33-37). Peaks are automatically collected and prepared for individual LC-MS runs. Although more tedious than a MudPIT, off-line chromatographic separations in general produce better chromatographic resolutions. In our opinion, off-line separations provide an excellent opportunity to reduce the sample complexity of mammalian protein extracts and further increase the number of detected proteins.

c) Gel-based separations. Gel electrophoresis has traditionally been a successful methodology for the separation of complex protein mixtures. Proteins are either separated by one-dimensional gel electrophoresis (1-DE), separating proteins based on their molecular weight or, if a higher resolution is desired, by two-dimensional gel electrophoresis (2-DE) (38, 39). In 2-DE, proteins are first separated based on their isoelectric point and, in the second dimension, based on their molecular weight. In either case, individual spots or bands of interest are excised, digested and analyzed by LC-MS.

Protein identification and validation

Independent of the mass spectrometer used or the separation strategy applied, the goal of large-scale proteomics is the identification of all proteins present in a given sample. Traditionally, protein identifications were achieved by manually assigning a peptide sequence to every recorded tandem mass spectra (*de novo* sequencing) (40, 41). Although, successfully applied for the identification of individual proteins, this method is highly dependent on the spectral quality and the expertise of the investigator. Importantly, modern proteomics projects can easily generate several hundred thousand MS/MS spectra, clearly limiting the use of *de novo* sequencing efforts. Luckily, several powerful algorithms exist to automatically correlate the recorded MS/MS spectra to extensive protein sequence databases. The two most common search algorithms used in proteomics laboratories are SEQUEST (42) and MASCOT (43). Basically, both algorithms correlate the peptide peaks from an experimentally recorded tandem mass spectrum with protein sequences in public databases. These algorithms assign several specific scores reflecting parameters such as spectral quality, difference between best hit and second best hit *etc.* for every putative peptide spectra (44, 45). These algorithms work exceptionally well to identify proteins present in a sample. Nevertheless, in large-scale proteomics experiments a large number of assigned spectra possess marginal to low scores, and caution should be taken to minimize the number of false-positive identifications. Several strategies have been proposed in recent years to objectively filter initial search results and minimize false-positive identifications without being too stringent against less abundant proteins (25, 46-50). Basically, these algorithms assign a statistical significance or probability score to every search result, allowing investigators to

objectively filter their results. This should be of great interest to every proteomics research group, especially if time-consuming and expensive back-up experiments are planned based on the initial proteomics screen. As a rule of thumb, the following simple rules are suggested to ensure high quality search results:

1. Only accept high scoring peptide identifications, optimally with a defined probability score.
2. Identifications based on a single peptide should be treated with caution (although one might be biased against low abundance and small proteins).
3. Perform searches against a protein database supplemented with a "decoy" database. The decoy database consists of every protein sequence in its inverted amino acid sequence. Matches against these "reverse proteins" can be considered false-positive. The search results can easily be filtered to achieve an acceptable false-positive rate (25, 46).

Bioinformatics

Global proteomics projects generate several thousand high confidence protein identifications providing a significant challenge for data management, data mining and visualization efforts. As our understanding of biology becomes ever more complicated, we have reached a point in time where the actual management of data is a major stumbling block to the interpretation of results from modern proteomics platforms. However, there are several useful and proven data analysis and data mining strategies currently in use in modern genomics and proteomics laboratories.

Data mining. In general, large-scale proteomics or genomics projects contain several different sample conditions, for example healthy samples and matching disease samples, or different developmental time-points. The aim is to globally detect differences in these sample pools, and to gain insights into the underlying biological mechanisms.

The process of finding useful patterns in data can be described in many ways, with terms such as data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing widely used (51). Data clustering can be defined as the global sorting of related proteomics data based on identified proteins and/or analyzed samples using diverse mathematical algorithms (52). It should be considered as a simple, first pass for the identification of intrinsic patterns within the analyzed data. Clustering is especially useful if a large dataset with many different samples or sample conditions is analyzed (53). Several software packages are freely available from the web. One of the most commonly used cluster programs is Cluster 3.0, which was developed by Hoon and colleagues, and will easily tackle most user requests (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/index.html) (54). The clustered data can then be graphically viewed in heat map format using TreeView (http://jtreeview.souceforgenet.net). TreeView is highly interactive, allowing users to filter and manipulate the clustered data to their individual liking. Importantly, selected clusters can be exported as text files for further analysis. An alternative approach to data analysis and visualization is presented in Figures 3 and 4, using a pseudo-color correlation matrix, and a binary tree-structured vector quantization (http://www.cs.toronto.edu/~juris/btsvq/ downloads.html) and self-organizing maps (55, 56). Further, once one has clinically-meaningful protein signatures decision-support systems may be applied to aid in sample classification, *i.e.*, class prediction (57, 58).

Data annotation. Protein annotation can be referred as the known or the predicted function of a protein within a cell, tissue or organism. Several web portals or software programs are available for the systematic annotation of proteins identified in large-scale proteomics projects.

The Gene Ontology Consortium. The Gene Ontology Consortium (GO) (http://www.geneontology.org) is a systematic effort to globally annotate proteins in most common organisms and model systems (59, 60). The GO database is designed in a tree-like structure with three main branches; *biological process, molecular function, cellular component*. Proteomics data can be mapped to GO terms (if available) using several publicly available programs such as GOminer (http://discover.nci.nih.gov/gominer/) (61, 62). By mapping identified proteins to available GO terms, investigators can rapidly obtain useful information on the known or predicted function of a large number of proteins. Importantly, programs like GOminer or MouseSpec (http:// tap.med.utoronto.ca/~mchow/mousespec/index_orig.html) (63) can calculate if certain GO terms are statistically enriched within a set of identified proteins. This feature can be particularly useful if combined with cluster analysis. For example, a set of identified proteins are first clustered and a cluster of particular interest, such as proteins up-regulated in the disease state, are exported and the significantly enriched GO terms calculated. In some cases, this might provide investigators with some clue as to the function of an unannotated protein within a cluster.

Subcellular localization and motifs databases. The knowledge of the subcellular localization of an unannotated protein, might allow for a first pass prediction of the protein's molecular function. For examples, unannotated proteins found in the nucleus probably have a biological function related to this organelle. As described above, subcellular fractionation by density gradient centrifugation combined with LC-MS is a powerful way to obtain global information on subcellular localization. Additionally, many algorithms exist to predict the subcellular localization of detected proteins.

PSORT II is a publicly available algorithm (http://psort.ims.u-tokyo.ac.jp/) accessible through a web browser

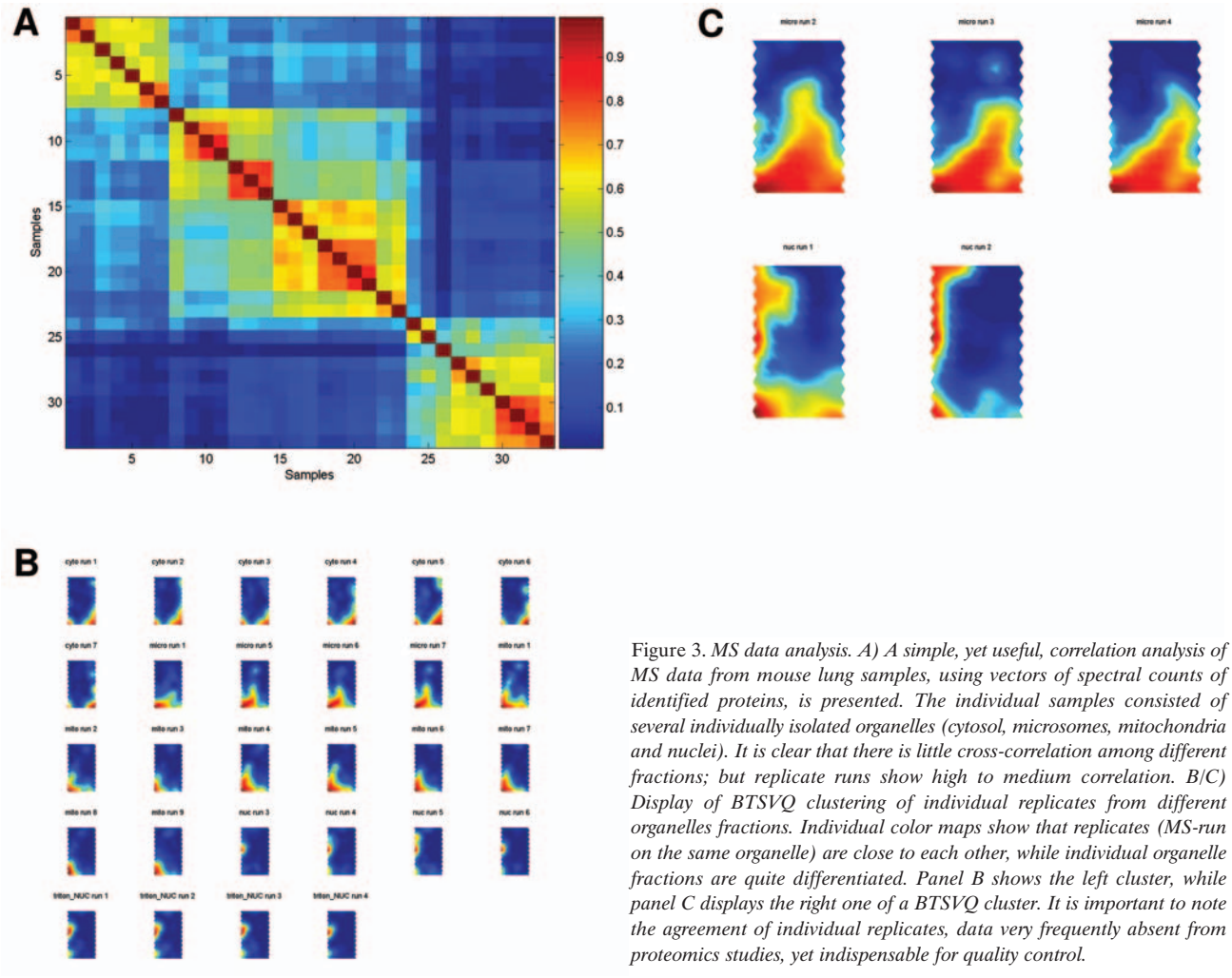

Figure 3. *MS data analysis. A) A simple, yet useful, correlation analysis of MS data from mouse lung samples, using vectors of spectral counts of identified proteins, is presented. The individual samples consisted of several individually isolated organelles (cytosol, microsomes, mitochondria and nuclei). It is clear that there is little cross-correlation among different fractions; but replicate runs show high to medium correlation. B/C) Display of BTSVQ clustering of individual replicates from different organelles fractions. Individual color maps show that replicates (MS-run on the same organelle) are close to each other, while individual organelle fractions are quite differentiated. Panel B shows the left cluster, while panel C displays the right one of a BTSVQ cluster. It is important to note the agreement of individual replicates, data very frequently absent from proteomics studies, yet indispensable for quality control.*

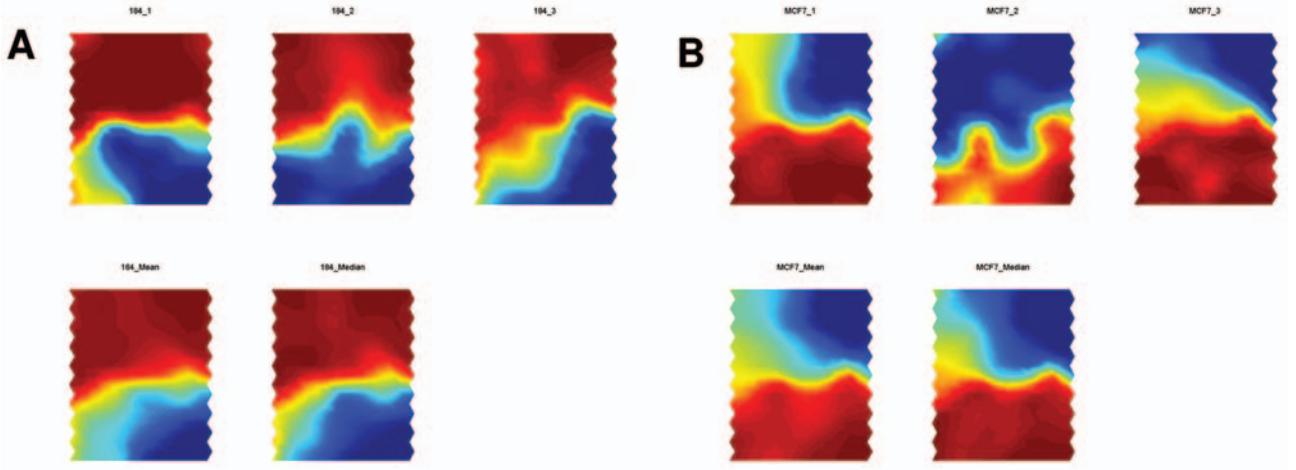

Figure 4. *Clustering and data analysis. MS analysis of two breast cancer cell lines from (179) using BTSVQ clustering. The first split separates 184 (Panel A) and MCF7 (Panel B) cell lines, using individual replicates, mean and median. Interestingly, in both cases, only two of the three replicates are highly similar, and dominate mean and median. However, the remaining single samples show the greatest difference between the cell lines.*

Figure 5. *Mapping MS-identified proteins to protein-protein interactions. MS profiles between two ovarian cell lines (ES2 and IOSE) were mapped into a protein-protein interactions database OPHID (75) (T.K and I.J., personal communications). The resulting network can be further analyzed using diverse approaches (180-182), such as hub identification and identification of subnetwork overlap. This last step is highlighted in this Figure, where black lines correspond to proteins and their neighborhood identified in both cell lines, green lines correspond to proteins and their neighborhood identified only in the ES2 cell line, while the red lines identify neighborhoods from the IOSE cell line. Further analysis and filtering would enable target selection for validation studies.*

(64). The algorithm predicts subcellular localization based on amino acid composition and the presence of targeting signals to various organelles. The amino acid sequence of a protein of interest is simply copied in FASTA format into the PSORT II search page to retrieve potential localization information. TargetP 1.1 is a similar web based algorithm (http://www.cbs.dtu.dk/services/TargetP/) capable of accepting amino acid sequences in FASTA format (65). The algorithm predicts the localization of eukaryotic proteins based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide, mitochondrial targeting peptide or secretory pathway signal peptide (SP). PSLT (Protein Subcellular Localization Tool) is a Bayesian network localization predictor that is based on the combinatorial presence of InterPro motifs and specific membrane domains (66). The algorithm generates a likelihood of localization to all organelles, and allows prediction of multicompartmental proteins. Importantly, similar algorithms (*e.g.*, TMpred; http://www.ch.embnet. org/software/TMPRED_form.html) exist for the prediction of transmembrane domains.

A protein domain is a portion of a protein with a defined tertiary structure. Domains often play a crucial role in the function of a protein, such as protein-protein biding, protein-DNA binding, kinase activity, *etc.* Knowledge of protein domains can provide clues about the protein function. The two most common and complete protein domain databases are Pfam (http://www.sanger.ac.uk/ Software/Pfam/) and InterPro (http://www.ebi.ac.uk/ interpro/) (67). Both databases can be reached through the ExPASY web portal (68). A global linkage of InterPro domains to proteins identified in a proteomics experiment on a database level can significantly speed up the progress.

Pathways. Proteins rarely function alone. In general, they form complexes (fine-tuned machineries), which coordinately carry out functions, summarized in *functional pathways*. The knowledge of significantly enriched, disturbed or up-regulated pathways, in a large-scale proteomics dataset can be highly informative, and could lead to a testable hypothesis. Several commercially, *e.g.*, Ingenuity Systems (http:// www.ingenuity. com/), or publicly available, *e.g.*, GenMAPP (http://www.genmapp.org/default.asp), KEGG (http:// www.genome.jp/ kegg), and BioCarta (http:// biocarta.com/genes) databases and browser software systems are available to map and organize proteins into predefined pathways (69-71). Basically, these programs possess interactive, user-friendly interfaces that allow visualizing proteomics and genomics data in the context of informative biological pathways. A recent addition to these tools – i-HOP – implements text mining for automating the process of protein annotation from PubMed (72).

Protein - protein interactions. Similarly to the pathway-based analysis described above, mapping MS-identified proteins into protein-protein interaction networks, available for multiple organisms in several databases BIND, DIP, GRID, HIMAP, HPRD, INTACT, MINT, MIPS, OPHID, POINT (73-82), will also enable the elucidation of important protein links. Graph analysis (*e.g.*, LEDA (83)), graph visualization software (*e.g.*, Cytoscape (84), MetNet3D (85), NAViGATOR (http://ophid. utoronto.ca/navigator/), WebInterViewer (86) and PIMWalker (87)) enable intuitive navigation through the networks of identified proteins. There are several portals to streamline access to the ever growing number of interaction databases, for example http://www.imb-jena.de/jcb/ppi/jcb_ppi_databases.html; http:// www.biopax.org/.

In Figure 5, a simplified example of both the power and complexity of such analyses, is presented, by mapping MS-identified protein from two ovarian cell lines into the OPHID protein interaction database (75).

Data resources. A wide variety of on-line databases and portals are available to handle data storage and interpretation. These resources cover repositories for MS data and supporting databases that aid in annotation and interpretation.

Standards. Standardizing data from proteomics experiments followed a path similar to the MIAME standard (88, 89). The MIAPE (Minimum Information About a Proteomics Experiment) standard defines the minimum reporting requirement for proteomics experiments, necessary for interpreting experimental results (90, 91). This standard was proposed by the HUPO's (Human Proteome Organization; http://www.hupo.org) PSI (Proteomics Standards Initiative; http://psidev.sourceforge.net/), combining early efforts from the Proteome Experimental Data Repository (PEDRo) (92, 93). To specifically handle PPI information, a data interchange standard has been developed (87, 90, 91, 94-96).

To handle MS data, the HUPO PSI General Proteomics Standards (GPS; http://psidev.sourceforge.net/gps/ index.html) has been extended to form the Proteomics Standards Initiative for Mass Spectrometry (PSI-MS; http://psidev.sourceforge.net/ ontology/index.html). The mzData standard captures peak-list data and aims to unite the large number of current formats (pkl's, dta's, mgf's, *etc.*). mzData is already supported by some software providers, while several other companies (*e.g.*, Bruker Daltonics, Kratos Analytical and Matrix Science) are currently implementing the mzData standard. The mzIdent standard captures parameters and results of search engines such as MASCOT and SEQUEST. Although the standards have been defined and accepted, the supporting ontology has not yet been created. The goal of the PSI Mass Spectrometry work group is to extend its coverage beyond the peak list to encompass the full chain from the mass spectrometer to the identification list.

Web-portals. The ExPASY web-portal (http://www.expasy.org) is a very useful starting point to retrieve information on proteins for all major model organisms (68, 97). In the start page, proteins are searched for by name, entry name or accession number. The investigator is directed to the next page which contains a multitude of useful information on this particular protein, such as literature references, a short description regarding the function, subcellular localization, tissue specificity, *etc.* (where available). Furthermore, direct links to other major information resources (domain databases, MGI, EMBL, *etc.*) are integrated. Additionally, the homepage of ExPASY provides the user with an exceptional selection of proteomics-related tools and software packages. Another useful resource for researchers working on mouse (*Mus musculus*) is Mouse Genome Informatics (MGI; http://www.informatics.jax.org/). Again, this resource provides literature references regarding the protein of interest and links to other major webpages. Additionally, MGI provides detailed genetic mappings, GO terms, expression data and phenotype information on knock-out mutants.

Mass spectrometry-based proteomics in medical research

The following paragraphs provide an overview of three major areas of medical proteomics (cancer, heart disease and diabetes). As a complete review of these broad research areas is beyond the scope of this article attention is focused on large-scale MS-based proteomics and 5-10 major publications in each category are reviewed. We apologize to investigators whose scientific accomplishments could not be reviewed in the following section. The overall characteristics of the dynamics and focus of the field are summarized in Figure 6. It is obvious that most of the applications of MS-based analysis have grown exponentially over the last 6 years.

Cancer research. Despite significant advanced in recent years, cancer remains a major public health risk. In the US alone, approximately half a million people die of cancer every year. Some of the major goals of cancer research remain the improvement of cancer therapy and the development of new diagnostic tools for early detection of the disease. Detailed information on the most common types of cancer and their associated statistics can be found on the web page of the National Cancer Institute (http://www.cancer.gov/). A large number of laboratories all over the world are now focusing on the application of proteomics to diverse kind of cancers.

a) Blood-based biomarker discovery. One of the major goals of medical proteomics is the detection of disease-specific proteins or protein signatures for diagnostic purposes. Ideally, these biomarkers could be used for screening and, thus, lead to early cancer detection, before any clinical symptoms are evident. Additionally, it would be ideal if biomarkers could be detected in a biological sample, such as blood, plasma or serum.

Blood-based samples are easily obtainable in the clinic, but several inherent complications have to be considered when dealing with plasma/serum samples if meaningful biological data is to be obtained. First, the collection of the plasma or serum samples and the downstream sample preparation protocol has to be standardized to ensure reproducible results. Second, the extreme complexity and skewed protein abundance (albumin and immunoglobulins make $\sim$70% of the total protein) render the analysis of plasma/serum samples extremely challenging. Third, improper experimental design can severely impede the results of the screen and their interpretation (98).

Despite these complications, a large number of investigators are applying proteomics technologies to blood-based samples with the hope of discovering meaningful, disease-specific biomarkers. Some of these efforts will be reviewed below.

SELDI-TOF-MS. SELDI-TOF-MS is still the most common proteomics platform for the analysis of serum (99-102). The original platform was introduced by Ciphergen Biosystems and is basically a low resolution time-of-flight mass spectrometer. The innovation applied by Ciphergen was the coating of the sample target with different chromatography resins (*e.g.*, ion exchange, reverse phase, metal ion binding, *etc.*), to reduce the initial sample complexity. The instrument records the m/z of ion species between 500-20,000 mass units. More recently, higher resolution systems, such as the hybrid quadrupole time-of-flight instruments (*e.g.*, QSTAR by Applied Biosystems) have been coupled to SELDI chips (103, 104). The individually recorded m/z values are normalized and analyzed by specific computer programs. In general, a digitized image of the recorded mass spectrum with the m/z value on the x-axis and a relative intensity value on the y-axis is generated for each sample. The samples are randomly divided into control and disease samples and highly sophisticated pattern recognition algorithms are applied to obtain reproducible differences between the two sample pools. The differences (*e.g.*, disease-specific pattern profiles) can be used to classify serum samples. To our knowledge, the first study using this technology to analyze the proteomic patterns in serum of ovarian cancer patients was published by the laboratory of Liotta at the National Institute of Health (NIH), USA, in 2002 (105). Although, initial attempts using this technology have produced promising results from a diagnostic point of view, one of the big drawbacks of this technology is the lack of unambiguous protein identifications based on molecular mass alone. This is a clear disadvantage of SELDI-based proteomics screens as only accurate knowledge of the disease-specific proteins could potentially result in the development of treatments or drugs. Improvements in MS technology (*e.g.*, higher resolution MS or top-down sequencing strategies) could potentially be an answer to this problem.

Importantly, the original data was subjected to multiple re-analyses (57, 98, 106-108), leading to a better understanding of the benefits and limitations of the markers and technology (109).

LC-MS-based screening of blood samples. To overcome the limitation of protein identification by SELDI-TOF-MS, several investigators have developed proteomics platforms based on microcapillary liquid chromatography coupled to mass spectrometers. Especially the laboratory of Smith from the Pacific Northwest National Laboratory has been active in analyzing the human plasma proteome by this method (110-116). A very interesting way of visualizing proteomics data obtained by LC-MS-based proteomics has been presented by the research groups of Aebersold and Emili. Basically, the generated data files are plotted in three dimensions using chromatographic retention time, m/z and ion intensity as the three dimensions. The generated peptide maps are compared between healthy and disease individuals to identify disease-specific peptides or peptide signatures (117, 118). We envision that the application and constant improvement of these technologies and software programs to the field of biomarker research will produce several interesting, putative target proteins for further clinical validation. Especially, the application of high resolution Fourier transform ion cyclotron mass spectrometers (FT-ICR-MS), capable of accurately recording peptide masses, should further accelerate this field of proteome research.

b) Laser capture microscopy. As proteomics is a highly technology-driven field of research, new developments are constantly being introduced to the research community. One of these new technologies is *laser capture microscopy*. This methodology is very useful when dealing with the problem of tissue heterogeneity (tissues are made up of many different cell types) (119). Laser capture microscopy was

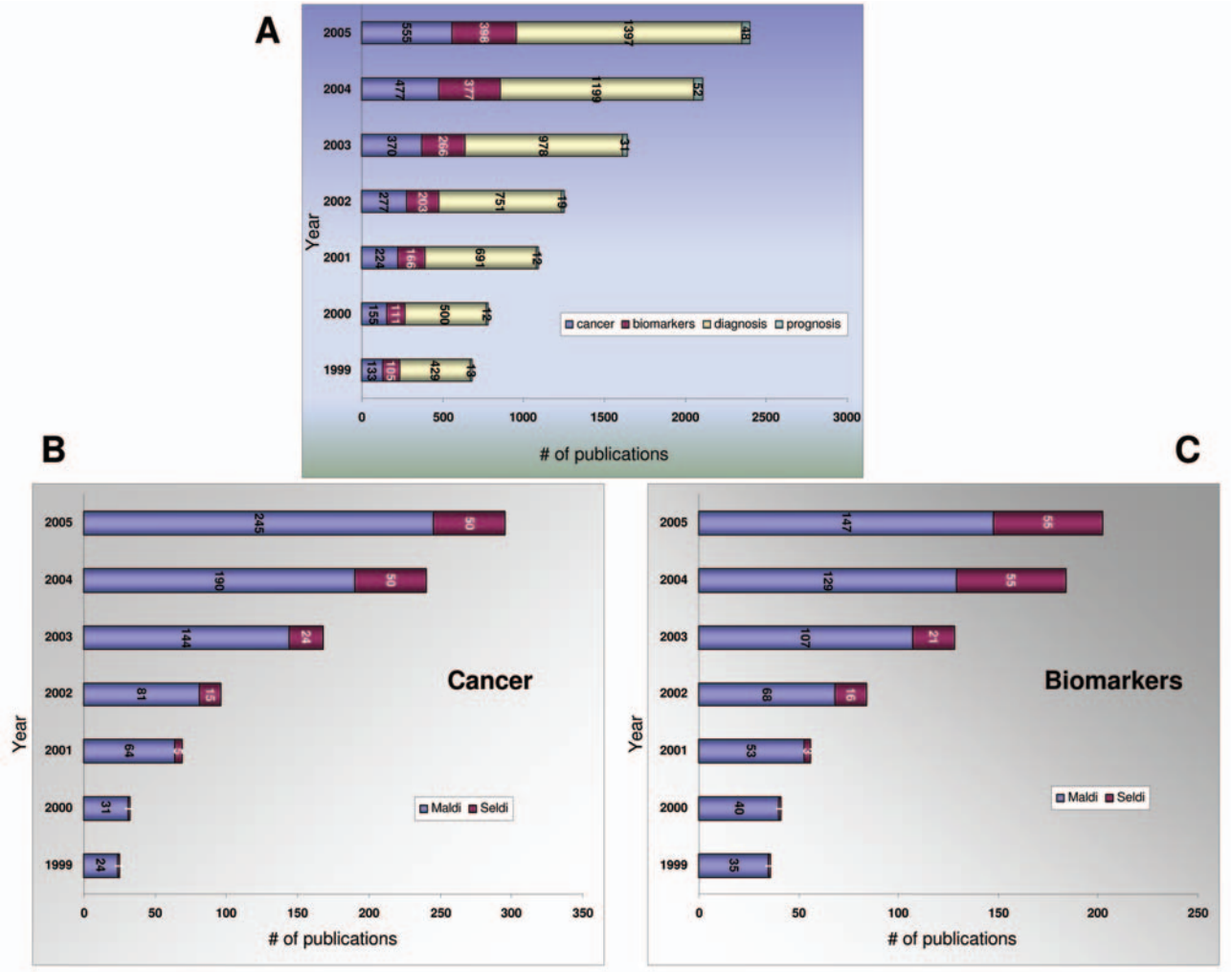

Figure 6. *Literature-based characterization of the MS field. A) Comparative differences of PubMed references for MS and cancer, biomarkers, diagnosis and prognosis over the last 6 years. Note that the 2005 numbers are extrapolated from the actual values for 2005 and expected growth predicted from the past. B) Differences of MALDI and SELDI for cancer-related research. C) Differences between MALDI and SELDI for biomarker identification and use.*

developed at the NIH and the first commercial instrument was available in 1997 (Arcturus Engineering). Basically, thin tissue sections (paraffin-embedded or frozen) are cut with a microtome and placed on a slide. A laser is used to fire at cells of interest and fuse them to an adhesive film, which can be removed to collect these cells. The collected cells are lysed and analyzed by various proteomics techniques. Several technical innovations have been introduced in recent years allowing for easy sample collection (*e.g.*, direct transfer of the captured cells into a sample collection tube) (119). Laser capture microscopy has several inherent disadvantages, which make its combination with current proteomics technologies difficult:

1. Freshly-frozen tissue should be used, as formalin-fixed tissue results in extensive protein crosslinkage.
2. Many of the routinely-used dyes for tissue staining interfere with proteomics technologies.
3. It is a particularly time-consuming procedure and will, therefore, most probably not be used for routine high-throughput proteomics research.

Despite all these problems, several applications of laser capture microscopy have been published in recent years. A study published by Mouledous *et al.* described a novel procedure for the analysis of brain tissue by navigated laser capture microscopy (120, 121). This procedure used fixed, unstained brain tissue sections, as staining was shown to result in problematic 2-DE. The authors were able to show that their methodology gave 2-DE results comparable to manually dissected brain tissue. Importantly, downstream MALDI-TOF-MS was not affected by this methodology and gave

similar protein identifications. An interesting study by Costello and colleagues combined laser capture microscopy and 2-DE to analyze differences in pancreatic ductal adenocarcinoma (122). First, laser capture microscopy was used to enrich for normal and malignant pancreatic ductal epithelial cells. Protein extracts were separated by 2-DE and the proteins were visualized by silver staining. Comparison of the 2 protein profiles showed 9 proteins that were differentially-regulated. One of the up-regulated proteins in the cancer samples was the calcium-binding S100A6 proteins. The expression pattern could be validated by immunohistochemical analysis in 46 pancreas cancer patients.

c) Imaging mass spectrometry and profiling of tissue sections. As described above, MALDI-TOF-MS has long been a useful tool for the detection of proteins. More recently, this MS technique has been applied to the direct analysis of tissue sections, mainly by the group of Caprioli at Vanderbilt University (123-125). One big advantage of this methodology is that less sample manipulation (*e.g.*, homogenization) is required and spatial tissue integrity is maintained. In general, frozen tissue sections (5-20 μm) are applied to stainless steel or glass sample slides and thawed directly on the target plate. Matrix solution is applied and mass spectral data is acquired from defined spots on the tissue section. One advantage of this type of mass spectrometry is that m/z values and spatial localization of distinct polypeptides are recorded within a tissue section (17). Although, the cellular localization of polypeptides may hold important diagnostic values, it has to be mentioned that individual proteins cannot be unambiguously identified by molecular weight alone. In a nice study published by Schwartz *et al.* (126), snap-frozen normal brain tissue was compared to brain tumor specimens. The polypeptide spectra recorded from the individual tissue section clearly allowed the authors to distinguish normal brain tissue from brain tumor tissue. Even tumor tissue of different histological grades could be distinguished by this methodology. In another study by Caprioli and colleagues imaging MALDI-MS was used to classify tumor subsets in non-small cell lung cancer (58). In this study, 79 lung tumors were compared to 14 normal lung tissues. Protein profiling directly on the tissue section was carried out as described and the recorded profiles were subject to detailed statistical analysis. The use of 82 differentially-expressed MS signals allowed the authors to distinguish (with 100% accuracy) between cancerous lung tissue and normal control samples. The model was equally efficient in classifying the different subsets of lung cancer. Importantly, this form of molecular profiling can also be used to classify tissue samples where histological classification might be difficult or impossible. In our opinion, it is going to be very interesting to follow the technical developments of this methodology over the next couple of years.

Several recent studies followed the early interest in proteomic profiling of ovarian cancer (127-129). Wang *et al.* (128) used a two-dimensional liquid phase separation method with non-porous reversed-phase high performance liquid chromatography to separate proteins from human ovarian epithelial whole cell lysates. MALDI-TOF was used to identify the protein expression in ovarian cancer cells *versus* non-neoplastic ovarian epithelial cells. The authors identified 300 novel proteins, which constituted ~60% of the detected proteins. Smith-Beckerman *et al.* (127) aimed at detecting proteins involved in pre-neoplastic changes in the ovarian surface epithelium, leading to overt malignancy, focusing on patients with a familial history of ovarian/breast cancer and mutations in the *BRCA1* tumor suppressor gene.

Heart disease. Heart failure presents the leading cause of morbidity and mortality worldwide (130). Although, not a uniform disease, there are many reasons for heart failure (hypertension, myocarditis, cardiomyopathy, *etc.*). Importantly, inadequate diagnostic measures exist to detect heart disease at an early stage and eventually cure or prevent the disease from progressing further. Proteomics research, in combination with genomics and physiological data, could be one methodology to bring new insight into cardiovascular research. Analyzing the heart proteome involves several analytical complications that could complicate the collection of meaningful biological data. The overall complexity and skewed composition of the heart proteome represent a serious difficulty even to modern mass spectrometers. Especially the presence of a few highly abundant housekeeping proteins, such as mitochondrial proteins, or structural components of the muscle fiber complicate the analysis of the heart proteome (131, 132). We have recently published a detailed article providing some solutions and optimized sample preparation protocols for the analysis of mouse heart tissue (18).

a) Proteomics analysis of healthy heart tissue. Several useful "resource type" proteome maps of healthy heart tissue have been published in recent years. These publications basically provide an inventory of proteins detectable by MS in heart extracts or specific organelles isolated from heart tissue. The initial attempts were done by 2-DE followed by MS of individual spots by the laboratories of Dunn and Jungblut (133, 134). In 2003, Taylor and colleagues published a detailed list of proteins identified in highly purified mitochondria isolated from the human heart. The authors were able to identify 615 proteins by a combination of sucrose gradient centrifugation, 1-DE and LC-MS (30). Many of the identified proteins were well-described members of mitochondrial multiprotein complexes, such as components of the oxidative phosphorylation machinery. Interestingly, approximately 19% of the identified proteins had no known or described biochemical roles within this organelle. Taylor *et al.* (30) provided, for the first time, a

detailed resource of a purified human organelle that resulted in the discovery of many novel putative mitochondrial proteins. This could significantly improve our understanding of the biochemical pathways and cellular functions of this important organelle. Recently our laboratory has published a comprehensive map of four organellar preparations from healthy adult mouse hearts (18). An optimized strategy for the isolation and solubilization of mitochondria (salt and detergent extraction), microsomes (mixed membrane preparation) and cytosol was developed. Each protein extract was digested with a combination of endoproteinase Lys-C and trypsin and the resulting peptide mixtures were analyzed by the MudPIT procedure. Approximately 400-600 unique proteins were detected in each fraction, from a total of 1,230 detected proteins (estimated false-positive rate <1%). This publication provides the most detailed proteome map of healthy mouse heart tissue and could provide new insights into the physiological processes controlling heart function. We are currently in the process of analyzing several established mouse models of heart disease (dilated cardiomyopathy, myocardial infarction, pressure overload, *etc.*) by proteomics technologies with the goal of gaining new insights into the mechanisms leading to these devastating abnormalities.

b) Proteomics analysis of heart disease. Several papers have been published applying proteomics technologies to study heart disease. Pan *et al.* reported the application of expression proteomics to the analysis of mouse hearts from hyper- and hypocontractile phospholamban mutants (26). Phospholamban (PLN), a membrane protein of the sarco(endo)plasmic reticulum, is an important regulator of cardiac contractility through its interaction and regulation of the sarco(endo)plasmic reticulum Ca^{2+} ATPase. In this study two different mouse strains were directly compared by MudPIT profiling. First, microsomal membranes were isolated by differential ultracentrifugation from PLN null mice (high cardiac contractility) and compared to transgenic mice overexpressing a superinhibitory PLN mutation (I40A). In total, 782 proteins were identified and it was demonstrated on a global scale that calcium dysregulation resulted in severe biochemical adaptations. A very similar study published by Kranias and colleagues employed 2-DE in combination with MS to analyze changes in the cardiac proteome of PLN knock-out mice and age-matched wild-type animals (135). Additionally, ^{32}P autoradiography was used to analyze changes in the phosphorylation status of the two mouse strains. Approximately 100 proteins with significant changes in expression level could be detected in the PLN knock-out mouse. Several changes in protein phosphorylation, suggesting important regulatory changes, were detected by autoradiography. These findings, which are in good agreement with the study of Pan *et al.*, suggest severe biochemical adaptations in the cardiac muscle proteome upon deletion of phospholamban. The group of Van Eyk has made extensive use of a transgenic mouse model of accelerated dilated cardiomyopathy, in which the small GTPase Rac1 is constitutively expressed in the myocardium (136, 137). In a study published in 2003 (136), these authors identified a significant increase of creatine kinase M-chain in the transgenic mouse model. Recently, the same group analyzed a cytosolic and myofilament-enriched extract from the same mouse model (137). Twelve proteins were identified as significantly changed in the cytosolic extract, including the tubulin beta chain, manganese superoxide dismutase and malate dehydrogenase. Interestingly, by analyzing different time-points throughout development of the disease, all three proteins are suggested to play a role in the early response to cardiac hypertrophic failure. Rose and colleagues used a 2-DE-based strategy to compare endomyocardial biopsies from patients rejecting their transplanted hearts and from patients showing no signs of rejection (138). Over 100 proteins were found to be up-regulated during rejection of the organ (2-fold to 50-fold). Using MS, 13 proteins were identified as either cardiac-specific or heat-shock proteins. Most interestingly, two of these proteins, alpha-B-crystallin and tropomyosin, were measured in the serum of patients following heart transplantation by ELISA technology. The mean levels of both proteins were significantly higher in patients showing signs of organ rejection.

Diabetes research. Diabetes, the disease in which the body does not produce or properly use insulin, affects a large number of people in the western, industrialized nations ($\sim$18 million people in the United States are diabetic) (American Diabetes Association; http://www.diabetes.org). The disease is associated with aggressive atherosclerosis and is emerging as the leading cause of end-stage renal failure. Although the pathological and pathophysiological indices of diabetic vascular complications are well-described, the precise molecular events have only recently been discovered (139-141). Global protein profiling in diabetes might reveal some of these molecular events and could lead to a better understanding of the disease.

a) Urinary proteomics in diabetes. Renal disease in patients with diabetes mellitus has become the leading cause of renal failure in the western world (142). Microalbuminuria is currently the only diagnostic test available for the early diagnosis of diabetic nephropathy. It would be greatly beneficial to identify other markers of renal dysfunction that could be used as biomarkers of diabetic nephropathy. Several papers have been published in recent years analyzing the urinary proteome, in the hope of identifying other biomarkers indicative of early diabetic renal damage. The study by Mischak *et al.* applied capillary electrophoreses coupled to MS (CE-MS) to generate a polypeptide map of human urine from 39 healthy and 112 type 2 diabetics (143). Interestingly,

the urinary peptide fingerprint differed significantly between the healthy and the diabetic patients and led to the establishment of a disease pattern. In patients with advanced renal disease (higher albuminuria), the peptide pattern was highly indicative of diabetic kidney damage. Using MS/MS, the authors were able to identify three polypeptides as fragments of the relaxin-like factor, Tamm-Horsfall protein (THP) and serum albumin. This could be considered a proof-of-principle experiment, as these proteins are known to be affected in kidney disease. The same group recently published a similar study profiling the urinary proteome of type 1 diabetics in comparison to age-matched control patients (144). The study identified more than 1,000 polypeptides in the urine and characterized them based on their m/z and retention time. Interestingly, 54 polypeptides were uniquely identified in type 1 diabetics, generating a significant set of putative, urinary disease biomarkers. It will be very interesting to see how this methodology performs on larger patient groups and how these preliminary biomarkers perform in such studies. Additionally, the validation of some putative biomarkers by alternative methodologies (*e.g.*, ELISA) would be highly desirable.

b) Protein profiling studies in diabetes. An elegant way to examine some of the cellular mechanisms that lead to diabetic vascular complications would be a global comparison of protein expression profiles of diabetic patients compared to age-matched, healthy individuals. Several excellent mouse models of diabetes (*e.g.*, db/db mouse) exist and could be used to monitor the global changes of protein expression under defined laboratory conditions (same genetic background, weight, age, *etc.*). These findings could be followed up in human patients. Unfortunately, to our knowledge, relatively few studies of this kind are to be found in the literature. A study by Klein and colleagues compared the whole-kidney protein expression patterns of the diabetic OVE26 mouse with the non-diabetic background strain by 2-DE (145). This study identified 41 proteins by MALDI-TOF-MS that were differentially-expressed in the diabetic animal. Among these were an elastase inhibitor and an elastase, resulting in the hypothesis of disturbed elastin expression in these diabetic animals. Western blotting and immunohistochemistry were able to validate this finding and clearly demonstrate the usefulness of expression proteomics in diabetic renal disease. As similar strategy was applied by Jiang *et al.* to study alterations in the proteome of red blood cell membranes in type 2 diabetes (146). Forty-two proteins could be identified that were differentially-expressed in the diabetic patients. Among those was flotillin-1, a recently discovered membrane protein of lipid rafts. This discovery could be potentially interesting as flotillin-1 has been linked to activation of the glucose transporter 4 in adipocytes in response to insulin. We expect a multitude of studies applying proteomics to investigate the cellular mechanisms leading to diabetic vascular damage in the next couple of years.

c) The use of MS to detect protein glycation. Protein glycation, the non-enzymatic reaction between reducing sugars and protein-bound amino groups results in the formation of a wide variety of structurally-different *Advanced Glycation Endproducts (AGEs)* (147, 148). AGEs have been implicated in the pathogenesis of diabetic complications (149). As protein glycation is a posttranslational protein modification, MS is an ideal tool for their detection. Several groups have used MS to detect protein glycation *in vitro* and *in vivo* (150-177). The use of MALDI-TOF-MS for the analysis of protein glycation was first reported by Lapolla and co-workers (176), who have since published a large number of manuscripts dealing with this subject (152, 154, 155, 157, 158, 161-165, 167, 169-177). Initially, they applied MALDI-TOF-MS to measure the protein glycation of BSA *in vitro*. They were able to detect an increase in the molecular weight of BSA dependent on the incubation time and the glucose concentration. After this initial success, the authors analyzed plasma proteins from diabetic and control patients and were able to detect differences in the molecular mass of the major plasma proteins (serum albumin and immunoglobulins). One drawback of MALDI-TOF-MS with intact proteins is the low resolution and mass accuracy in the mass range of intact proteins (>25 kDa). Therefore, the exact mass increase and/or the modification site on the protein could not be detected in these studies. In 2002, Humeny *et al.* reported the use of MALDI-TOF-MS in combination with enzymatic digestion for the determination of specific protein glycation products *in vitro* (168). By digesting glycated lysozyme with endoproteinase Glu-C and comparing the resulting peptide fingerprints to native lysozyme, the authors were able to identify several structural specific protein glycation products. Furthermore, the exact sites of the modifications could be precisely determined. The same group also demonstrated the ability of this methodology for the relative quantification of specific glycation products, produced upon incubation of lysozyme with different sugar concentrations *in vitro* (160). Likewise, the group of Thornalley and co-workers have clearly demonstrated the use of LC-MS-based methodologies to quantitatively screen for AGEs in tissue and plasma proteins. They measured the concentrations of 12 different AGEs by LC-MS and could clearly show that protein glycation was increased in the plasma and at sites of vascular complications (renal glomeruli, retina and peripheral nerve) in diabetes (159). Finally, the same group used a peptide mapping strategy to detect hotspots of protein modification on human serum albumin modified with methylglyoxal (178). Interestingly, the major site of protein modification was located at the drug-binding site II and the active site of albumin-associated esterase activity (Arginine 410). This publication demonstrated that protein glycation can clearly result in functional disruptions of the modified protein.

Conclusion

In conclusion, an overview of the current developments in the field of medical proteomics and bioinformatics has been provided, including current MS technologies, sample separation and preparation strategies used in modern proteomics laboratories. Some of the proven bioinformatics and data mining strategies are described throughout this article, and should provide the readers with valuable ideas on how to organize and analyze large datasets produced in proteomics projects. Data mining seems to be a considerable bottleneck in large-scale systems biology and obtaining biologically useful data has to be of the highest priority. Finally, some of the current advances in three major areas of applied medical proteomics were reviewed: cancer proteomics, proteomics of heart tissue/disease and proteomics in diabetes research. We expect that constant improvement in MS hardware, bioinformatics tools and sample preparation strategies, in combination with well-designed biological experiments, will produce many exciting results over the next decade.

References

1 Steen H and Mann M: The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol *5(9)*: 699-711, 2004.

2 Aebersold R and Mann M: Mass spectrometry-based proteomics. Nature *422(6928)*: 198-207, 2003.

3 Gibbs RA *et al*: Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature *428(6982)*: 493-521, 2004.

4 Gregory SG *et al*: A physical map of the mouse genome. Nature *418(6899)*: 743-750, 2002.

5 Venter JC *et al*: The sequence of the human genome. Science *291(5507)*: 1304-1351, 2001.

6 Waterston RH *et al*: Initial sequencing and comparative analysis of the mouse genome. Nature *420(6915)*: 520-562, 2002.

7 Lander ES *et al*: Initial sequencing and analysis of the human genome. Nature *409(6822)*: 860-921, 2001.

8 Fenn JB *et al*: Electrospray ionization for mass spectrometry of large biomolecules. Science *246(4926)*: 64-71, 1989.

9 Karas M and Hillenkamp F: Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60(20)*: 2299-2301, 1988.

10 Gavin AC *et al*: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415(6868)*: 141-147, 2002.

11 Seibert V *et al*: Surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI TOF-MS) and ProteinChip technology in proteomics research. Pathol Res Pract *200(2)*: 83-94, 2004.

12 Tang N, Tornatore P and Weinberger SR: Current developments in SELDI affinity technology. Mass Spectrom Rev *23(1)*: 34-44, 2004.

13 Thiede B *et al*: Peptide mass fingerprinting. Methods *35(3)*: 237-247 2005.

14 Simpson R: Proteins and Proteomics: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 2002.

15 Stuhler K and Meyer HE: MALDI: more than peptide mass fingerprints. Curr Opin Mol Ther *6(3)*: 239-248, 2004.

16 James P: Proteome Research: Mass Spectrometry. Springer-Verlag, Berlin, Heidelberg, New York, 2001.

17 Reyzer ML and Caprioli RM: MALDI mass spectrometry for direct tissue analysis: a new tool for biomarker discovery. J Proteome Res *4(4)*: 1138-1142, 2005.

18 Kislinger T *et al*: Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. J Am Soc Mass Spectrom *16(8)*: 1207-1220, 2005.

19 Liu H, Sadygov RG and Yates JR III: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76(14)*: 4193-4201, 2004.

20 Washburn MP, Wolters D and Yates JR III: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol *19(3)*: 242-247, 2001.

21 Link AJ *et al*: Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol *17(7)*: 676-682, 1999.

22 Kislinger T and Emili A: Multidimensional protein identification technology: current status and future prospects. Expert Rev Proteomics *2(1)*: 27-39, 2005.

23 Delahunty C and Yates JR III: Protein identification using 2D-LC-MS/MS. Methods *35(3)*: 248-255, 2005.

24 Kislinger T *et al*: Proteome dynamics during C2C12 myoblast differentiation. Mol Cell Proteomics *4(7)*: 887-901, 2005.

25 Kislinger T *et al*: PRISM, a generic large scale proteomic investigation strategy for mammals. Mol Cell Proteomics *2(2)*: 96-106, 2003.

26 Pan Y *et al*: Identification of biochemical adaptations in hyper- or hypocontractile hearts from phospholamban mutant mice by expression proteomics. Proc Natl Acad Sci USA *101(8)*: 2241-2246, 2004.

27 Schirmer EC *et al*: Identification of novel integral membrane proteins of the nuclear envelope with potential disease links using subtractive proteomics. Novartis Found Symp *264*: 63-76; discussion 76-80, 227-30, 2005.

28 Durr E *et al*: Direct proteomic mapping of the lung microvascular endothelial cell surface *in vivo* and in cell culture. Nat Biotechnol *22(8)*: 985-992, 2004.

29 Yates JR *et al*: Proteomics of organelles and large cellular structures. Nature Rev Mol Cell Biol *6*: 702-714, 2005.

30 Taylor SW *et al*: Characterization of the human heart mitochondrial proteome. Nat Biotechnol *21(3)*: 281-286, 2003.

31 Andersen JS *et al*: Directed proteomic analysis of the human nucleolus. Curr Biol *12(1)*: 1-11, 2002.

32 Andersen JS *et al*: Nucleolar proteome dynamics. Nature *433(7021)*: 77-83, 2005.

33 Wang Y *et al*: Integrated capillary isoelectric focusing/nano-reversed phase liquid chromatography coupled with ESI-MS for characterization of intact yeast proteins. J Proteome Res *4(1)*: 36-42, 2005.

34 Wang Y *et al*: Effects of chromatography conditions on intact protein separations for top-down proteomics. J Chromatogr A *1073(1-2)*: 35-41, 2005.

35 Li X *et al*: Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project. Proteomics *5(13)*: 3423-3441, 2005.

36 Reschiglian P *et al*: Field-flow fractionation and biotechnology. Trends Biotechnol *23(9)*: 475-483, 2005.

37 Sheng S, Chen D and Van Eyk JE: Multi-dimensional liquid chromatography separation of intact proteins by chromatographic focusing and reversed phased of the human serum proteome: optimization and protein database. Mol Cell Proteomics, 2005.

38 Klose J: Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik *26(3)*: 231-243, 1975.

39 O'Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250(10)*: 4007-4021, 1975.

40 Gevaert K and Vandekerckhove J: Protein identification methods in proteomics. Electrophoresis *21(6)*: 1145-1154, 2000.

41 Shevchenko A *et al*: Rapid '*de novo*' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. Rapid Commun Mass Spectrom *11(9)*: 1015-1024, 1997.

42 Eng JK, McCormack AL and Yates JRI: An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J Am Soc Mass Spectrom *11*: 976-989, 1994.

43 Perkins DN *et al*: Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20(18)*: 3551-3567, 1999.

44 Elias JE *et al*: Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods *2(9)*: 667-675, 2005.

45 Kapp EA *et al*: An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. Proteomics *5(13)*: 3475-3490, 2005.

46 Peng J *et al*: Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res *2(1)*: 43-50, 2003.

47 Moore RE, Young MK and Lee TD: Qscore: an algorithm for evaluating SEQUEST database search results. J Am Soc Mass Spectrom *13(4)*: 378-386, 2002.

48 MacCoss MJ, Wu CC and Yates JR III: Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal Chem *74(21)*: 5593-5599, 2002.

49 Keller A *et al*: Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74(20)*: 5383-5392, 2002.

50 Lopez-Ferrer D *et al*: Statistical model for large-scale peptide identification in databases from tandem mass spectra using SEQUEST. Anal Chem *76(23)*: 6853-6860, 2004.

51 Jurisica I and Wigle DA: Knowledge Discovery in Proteomics. Chapman & Hall/CRC Press, 2006.

52 Quackenbush J: Computational analysis of microarray data. Nat Rev Genet *2(6)*: 418-427, 2001.

53 Cox B, Kislinger T and Emili A: Integrating gene and protein expression data: pattern analysis and profile mining. Methods *35(3)*: 303-314, 2005.

54 Eisen MB *et al*: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA *95(25)*: 14863-14868, 1998.

55 Sultan M *et al*: Binary tree-structured vector quantization approach to clustering and visualizing microarray data. Bioinformatics *18 Suppl 1*: S111-S119, 2002.

56 Kohonen T: Self Organizing Maps. Springer-Verlag, Berlin, 1995.

57 Arshadi N and Jurisica I: Integrating case-based reasoning systems with data mining techniques for discovering and using disease biomarkers. IEEE Transactions on Knowledge and Data Engineering. Special Issue-Mining Biological Data *17(8)*: 1127-1137, 2005.

58 Yanagisawa K *et al*: Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet *362(9382)*: 433-439, 2003.

59 Harris MA *et al*: The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res *32(Database issue)*: D258-261, 2004.

60 Ashburner M *et al*: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25(1)*: 25-29, 2000.

61 Zeeberg BR *et al*: GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol *4(4)*: R28, 2003.

62 Zeeberg BR *et al*: High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of common variable immune deficiency (CVID). BMC Bioinformatics *6*: 168, 2005.

63 Robinson MD *et al*: FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics *3(1)*: 35, 2002.

64 Nakai K and Horton P: PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci *24(1)*: 34-36, 1999.

65 Emanuelsson O *et al*: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol *300(4)*: 1005-1016, 2000.

66 Scott MS, Thomas DY and Hallett MT: Predicting subcellular localization *via* protein motif co-occurrence. Genome Res *14(10A)*: 1957-1966, 2004.

67 Mulder NJ *et al*: InterPro, progress and status in 2005. Nucleic Acids Res *33(Database Issue)*: D201-205, 2005.

68 Gasteiger E *et al*: ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res *31(13)*: 3784-3788, 2003.

69 Mlecnik B *et al*: PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. Nucleic Acids Res *33(Web Server issue)*: W633-637, 2005.

70 Dahlquist KD *et al*: GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. Nat Genet *31(1)*: 19-20, 2002.

71 Kanehisa M *et al*: The KEGG databases at GenomeNet. Nucleic Acids Res *30(1)*: 42-46, 2002.

72 Hoffmann R and Valencia A: Implementing the iHOP concept for navigation of biomedical literature. Bioinformatics *21 Suppl 2*: ii252-ii258, 2005.

73 Bader GD, Betel D and Hogue CW: BIND: the biomolecular interaction network database. Nucleic Acids Res *31(1)*: 248-250, 2003.

74 Breitkreutz BJ, Stark C and Tyers M: The GRID: the general repository for interaction datasets. Genome Biol *4(3)*: R23, 2003.

75 Brown KR and Jurisica I: Online predicted human interaction database. Bioinformatics *21(9)*: 2076-2082, 2005.

76 Han K *et al*: HPID: the human protein interaction database. Bioinformatics *20(15)*: 2466-2470, 2004.

77 Hermjakob H *et al*: IntAct: an open source molecular interaction database. Nucleic Acids Res *32(Database issue)*: D452-455, 2004.

78 Huang TW *et al*: POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. Bioinformatics, 2005 (in press).

79 Mewes HW *et al*: MIPS: a database for genomes and protein sequences. Nucleic Acids Res *30(1)*: 31-34, 2002.

80 Peri S *et al*: Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res *13(10)*: 2363-2371, 2003.

81 Xenarios I *et al*: DIP: the database of interacting proteins. Nucleic Acids Res *28(1)*: 289-291, 2000.

82 Zanzoni A *et al*: MINT: a Molecular INTeraction database. FEBS Lett *513(1)*: 135-140, 2002.

83 Mehlhorn K and Naher S: Leda: A Platform for Combinatorial and Geometric Computing. Cambridge University Press, 1999.

84 Shannon P *et al*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13(11)*: 2498-2504, 2003.

85 Yang Y *et al*: Integration of metabolic networks and gene expression in virtual reality. Bioinformatics *21(18)*: 3645-3650, 2005.

86 Han K, Ju BH and Jung H: WebInterViewer: visualizing and analyzing molecular interaction networks. Nucleic Acids Res *32(Web Server issue)*: W89-95, 2004.

87 Meil A, Durand P and Wojcik J: PIMWalker: visualising protein interaction networks using the HUPO PSI molecular interaction format. Appl Bioinformatics *4(2)*: 137-139, 2005.

88 Brazma A *et al*: Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. Nat Genet *29(4)*: 365-371, 2001.

89 Knudsen TB and Daston GP: MIAME guidelines. Reprod Toxicol *19(3)*: 263, 2005.

90 Orchard S *et al*: Further steps in standardisation. Report of the second annual Proteomics Standards Initiative Spring Workshop (Siena, Italy, April 17-20, 2005). Proteomics *5(14)*: 3552-3555, 2005.

91 Orchard S *et al*: Common interchange standards for proteomics data: public availability of tools and schema. Proteomics *4(2)*: 490-491, 2004.

92 Garwood K *et al*: PEDRo: a database for storing, searching and disseminating experimental proteomics data. BMC Genomics *5(1)*: 68, 2004.

93 Taylor CF *et al*: A systematic approach to modeling, capturing, and disseminating proteomics experimental data. Nat Biotechnol *21(3)*: 247-254, 2003.

94 Orchard S *et al*: Human Proteome Organisation Proteomics Standards Initiative Pre-congress Initiative. Proteomics, 2005.

95 Hermjakob, H *et al*: The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. Nat Biotechnol *22(2)*: 177-183, 2004.

96 Orchard S, Hermjakob H and Apweiler R: The proteomics standards initiative. Proteomics *3(7)*: 1374-1376, 2003.

97 Schneider M, Tognolli M and Bairoch A: The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. Plant Physiol Biochem *42(12)*: 1013-1021, 2004.

98 Baggerly KA *et al*: Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst *97(4)*: 307-309, 2005.

99 Xiao Z *et al*: Proteomic patterns: their potential for disease diagnosis. Mol Cell Endocrinol *230(1-2)*: 95-106, 2005.

100 Reddy G and Dalmasso EA: SELDI ProteinChip(R) Array Technology: protein-based predictive medicine and drug discovery applications. J Biomed Biotechnol *2003(4)*: 237-241, 2003.

101 Simpkins F *et al*: SELDI-TOF mass spectrometry for cancer biomarker discovery and serum proteomic diagnostics. Pharmacogenomics *6(6)*: 647-653, 2005.

102 Petricoin EF and Liotta LA: SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotechnol *15(1)*: 24-30, 2004.

103 Conrads TP *et al*: High-resolution serum proteomic features for ovarian cancer detection. Endocr Relat Cancer *11(2)*: 163-178, 2004.

104 Conrads TP *et al*: Cancer diagnosis using proteomic patterns. Expert Rev Mol Diagn *3(4)*: 411-420, 2003.

105 Petricoin EF *et al*: Use of proteomic patterns in serum to identify ovarian cancer. Lancet *359(9306)*: 572-577, 2002.

106 Baggerly KA, Morris JS and Coombes KR: Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics *20(5)*: 777-785, 2004.

107 Zhu W *et al*: Detection of cancer-specific markers amid massive mass spectral data. Proc Natl Acad Sci USA *100(25)*: 14666-14671, 2003.

108 Sorace JM and Zhan M: A data review and re-assessment of ovarian cancer serum proteomic profiling. BMC Bioinformatics *4*: 24, 2003.

109 Hu J *et al*: The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. Brief Funct Genomic Proteomic *3(4)*: 322-331, 2005.

110 Qian WJ *et al*: Quantitative proteome analysis of human plasma following *in vivo* lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. Mol Cell Proteomics *4(5)*: 700-709, 2005.

111 Qian WJ *et al*: Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res *4(1)*: 53-62, 2005.

112 Qian WJ *et al*: Comparative proteome analyses of human plasma following *in vivo* lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. Proteomics *5(2)*: 572-584, 2005.

113 Jacobs JM *et al*: Utilizing human blood plasma for proteomic biomarker discovery. J Proteome Res *4(4)*: 1073-1085, 2005.

114 Shen Y *et al*: Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. Anal Chem *76(4)*: 1134-1144, 2004.

115 Shen Y *et al*: Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS. Anal Chem *76(1)*: 144-154, 2004.

116 Shen Y *et al*: Characterization of the human blood plasma proteome. Proteomics *5(15)*: 4034-4045, 2005.

117 Li XJ *et al*: A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. Anal Chem *76(13)*: 3856-3860, 2004.

118 Radulovic D *et al*: Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics *3(10)*: 984-997, 2004.

119 Kunz GM Jr and Chan DW: The use of laser capture microscopy in proteomics research – a review. Dis Markers *20(3)*: 155-160, 2004.

120 Mouledous L *et al*: Navigated laser capture microdissection as an alternative to direct histological staining for proteomic analysis of brain samples. Proteomics *3(5)*: 610-615, 2003.

121 Mouledous L *et al*: Proteomic analysis of immunostained, laser-capture microdissected brain samples. Electrophoresis *24(1-2)*: 296-302, 2003.

122 Shekouh AR *et al*: Application of laser capture microdissection combined with two-dimensional electrophoresis for the discovery of differentially regulated proteins in pancreatic ductal adenocarcinoma. Proteomics *3(10)*: 1988-2001, 2003.

123 Caldwell RL and Caprioli RM: Tissue profiling by mass spectrometry: a review of methodology and applications. Mol Cell Proteomics *4(4)*: 394-401, 2005.

124 Chaurand P *et al*: Imaging mass spectrometry: principles and potentials. Toxicol Pathol *33(1)*: 92-101, 2005.

125 Crecelius AC *et al*: Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry. J Am Soc Mass Spectrom *16(7)*: 1093-1099, 2005.

126 Schwartz SA *et al*: Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. Clin Cancer Res *10(3)*: 981-987, 2004.

127 Smith-Beckerman DM *et al*: Proteome changes in ovarian epithelial cells derived from women with BRCA1 mutations and family histories of cancer. Mol Cell Proteomics *4(2)*: 156-168, 2005.

128 Wang H *et al*: Comprehensive proteome analysis of ovarian cancers using liquid phase separation, mass mapping and tandem mass spectrometry: a strategy for identification of candidate cancer biomarkers. Proteomics *4(8)*: 2476-2495, 2004.

129 Woong-Shick A *et al*: Identification of hemoglobin-alpha and -beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. Cancer Sci *96(3)*: 197-201, 2005.

130 Lee DS *et al*: Trends in heart failure outcomes and pharmacotherapy: 1992 to 2000. Am J Med *116(9)*: 581-589, 2004.

131 Arrell DK, Neverova I and Van Eyk JE: Cardiovascular proteomics: evolution and potential. Circ Res *88(8)*: 763-773, 2001.

132 Jager D, Jungblut PR and Muller-Werdan U: Separation and identification of human heart proteins. J Chromatogr B Analyt Technol Biomed Life Sci *771(1-2)*: 131-153, 2002.

133 Jungblut P *et al*: Protein composition of the human heart: the construction of a myocardial two-dimensional electrophoresis database. Electrophoresis *15(5)*: 685-707, 1994.

134 Baker CS *et al*: A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting. Electrophoresis *13(9-10)*: 723-726, 1992.

135 Chu G *et al*: Proteomic analysis of hyperdynamic mouse hearts with enhanced sarcoplasmic reticulum calcium cycling. FASEB J *18(14)*: 1725-1727, 2004.

136 Buscemi N *et al*: Proteomic analysis of Rac1 transgenic mice displaying dilated cardiomyopathy reveals an increase in creatine kinase M-chain protein abundance. Mol Cell Biochem *251(1-2)*: 145-151, 2003.

137 Buscemi N *et al*: Myocardial subproteomic analysis of a constitutively active Rac1 expressing transgenic mouse with lethal myocardial hypertrophy. Am J Physiol Heart Circ Physiol, 2005.

138 Borozdenkova S *et al*: Use of proteomics to discover novel markers of cardiac allograft rejection. J Proteome Res *3(2)*: 282-288, 2004.

139 Naka Y *et al*: RAGE axis: animal models and novel insights into the vascular complications of diabetes. Arterioscler Thromb Vasc Biol *24(8)*: 1342-1349, 2004.

140 Yan SF *et al*: Glycation, inflammation, and RAGE: a scaffold for the macrovascular complications of diabetes and beyond. Circ Res *93(12)*: 1159-1169, 2003.

141 Hudson BI *et al*: Blockade of receptor for advanced glycation endproducts: a new target for therapeutic intervention in diabetic complications and inflammatory disorders. Arch Biochem Biophys *419(1)*: 80-88, 2003.

142 Rabelink TJ, Stroes ES and Koomans HA: Mechanisms of cardiovascular injury in renal disease. Blood Purif *14(1)*: 67-74, 1996.

143 Mischak H *et al*: Proteomic analysis for the assessment of diabetic renal damage in humans. Clin Sci (Lond) *107(5)*: 485-495, 2004.

144 Meier M *et al*: Identification of urinary protein pattern in type 1 diabetic adolescents with early diabetic nephropathy by a novel combined proteome analysis. J Diabetes Complications *19(4)*: 223-232, 2005.

145 Thongboonkerd V *et al*: Alterations in the renal elastin-elastase system in type 1 diabetic nephropathy identified by proteomic analysis. J Am Soc Nephrol *15(3)*: 650-662, 2004.

146 Jiang M *et al*: Protein disregulation in red blood cell membranes of type 2 diabetic patients. Biochem Biophys Res Commun *309(1)*: 196-200, 2003.

147 Stitt AW, Frizzell N and Thorpe SR: Advanced glycation and advanced lipoxidation: possible role in initiation and progression of diabetic retinopathy. Curr Pharm Des *10(27)*: 3349-3360, 2004.

148 Ahmed N: Advanced glycation endproducts – role in pathology of diabetic complications. Diabetes Res Clin Pract *67(1)*: 3-21, 2005.

149 Jandeleit-Dahm KA, Lassila M and Allen TJ: Advanced glycation end products in diabetes-associated atherosclerosis and renal disease: interventional studies. Ann NY Acad Sci *1043*: 759-766, 2005.

150 Schmitt A, Gasic-Milenkovic J and Schmitt J: Characterization of advanced glycation end products: mass changes in correlation to side chain modifications. Anal Biochem, 2005.

151 Farah MA *et al*: Analysis of glycated insulin by MALDI-TOF mass spectrometry. Biochim Biophys Acta, 2005.

152 Lapolla A *et al*: Advanced glycation end products/peptides: an *in vivo* investigation. Ann NY Acad Sci *1043*: 267-275, 2005.

153 Kislinger T *et al*: Analysis of protein glycation products by MALDI-TOF/MS. Ann NY Acad Sci *1043*: 249-259, 2005.

154 Lapolla A *et al*: A matrix-assisted laser desorption/ionization mass spectrometry study of the non-enzymatic glycation products of human globins in diabetes. Rapid Commun Mass Spectrom *19(2)*: 162-168, 2005.

155 Lapolla A *et al*: The complexity of non-enzymatic glycation product sets of human globins. Diabetologia *47(10)*: 1712-1715, 2004.

156 Kislinger T, Humeny A and Pischetsrieder M: Analysis of protein glycation products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Curr Med Chem *11(16)*: 2185-2193, 2004.
157 Lapolla A *et al*: Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. J Am Soc Mass Spectrom *15(4)*: 496-509, 2004.
158 Lapolla A *et al*: Glyoxal and methylglyoxal levels in diabetic patients: quantitative determination by a new GC/MS method. Clin Chem Lab Med *41(9)*: 1166-1173, 2003.
159 Thornalley PJ *et al*: Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. Biochem J *375(Pt 3)*: 581-592, 2003.
160 Kislinger T *et al*: Relative quantification of N(epsilon)-(carboxymethyl)lysine, imidazolone A, and the Amadori product in glycated lysozyme by MALDI-TOF mass spectrometry. J Agric Food Chem *51(1)*: 51-57, 2003.
161 Lapolla A *et al*: Advanced glycation end-products/peptides: a preliminary investigation by LC and LC/MS. Farmaco *57(10)*: 845-852, 2002.
162 Lapolla A *et al*: Non-enzymatic glycation of IgG: an *in vivo* study. Horm Metab Res *34(5)*: 260-264, 2002.
163 Lapolla A *et al*: Advanced glycation end products: a highly complex set of biologically relevant compounds detected by mass spectrometry. J Mass Spectrom *36(4)*: 370-378, 2001.
164 Lapolla A, Fedele D and Traldi P: Diabetes and mass spectrometry. Diabetes Metab Res Rev *17(2)*: 99-112, 2001.
165 Lapolla A, Fedele D and Traldi P: The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes. Mass Spectrom Rev *19(5)*: 279-304, 2000.
166 Thornalley PJ *et al*: Mass spectrometric monitoring of albumin in uremia. Kidney Int *58(5)*: 2228-2234, 2000.
167 Lapolla A *et al*: Matrix-assisted laser desorption/ionization mass spectrometry, enzymatic digestion, and molecular modeling in the study of nonenzymatic glycation of IgG. J Am Soc Mass Spectrom *11(2)*: 153-159, 2000.
168 Humeny A *et al*: Qualitative determination of specific protein glycation products by matrix-assisted laser desorption/ionization mass spectrometry peptide mapping. J Agric Food Chem *50(7)*: 2153-2160, 2002.
169 Lapolla A *et al*: Direct evaluation of glycated and glyco-oxidized globins by matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom *13(1)*: 8-14, 1999.
170 Lapolla A *et al*: Evaluation of IgG glycation levels by matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom *11(12)*: 1342-1346, 1997.
171 Lapolla A *et al*: A highly specific method for the characterization of glycation and glyco-oxidation products of globins. Rapid Commun Mass Spectrom *11(6)*: 613-617, 1997.
172 Lapolla, A *et al*: The *in vitro* glycation of lysozyme and the influence of buffer concentration investigated by mass spectrometry. Rapid Commun Mass Spectrom *10(12)*: 1512-1518, 1996.
173 Lapolla, A *et al*: The *in vivo* glyco-oxidation of alpha- and beta-globins investigated by matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom *10(9)*: 1133-1135, 1996.
174 Lapolla A *et al*: An electrospray investigation on *in vitro* glycation of ribonuclease. Rapid Commun Mass Spectrom *10(2)*: 178-182, 1996.
175 Lapolla A *et al*: Matrix-assisted laser desorption/ionization capabilities in the study of non-enzymatic protein glycation. Rapid Commun Mass Spectrom *8(8)*: 645-652, 1994.
176 Lapolla A *et al*: A study on *in vitro* glycation processes by matrix-assisted laser desorption ionization mass spectrometry. Biochim Biophys Acta *1225(1)*: 33-38, 1993.
177 Lapolla A *et al*: Mass spectrometry in the study of advanced glycation processes, responsible for long-term diabetes complications. Rapid Commun Mass Spectrom *5(11)*: 527-533, 1991.
178 Ahmed N *et al*: Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. J Biol Chem *280(7)*: 5724-5732, 2005.
179 Sandhu C *et al*: Global protein shotgun expression profiling of proliferating mcf-7 breast cancer cells. J Proteome Res *4(3)*: 674-689, 2005.
180 King AD, Przulj N and Jurisica I: Protein complex prediction *via* cost-based clustering. Bioinformatics *20(17)*: 3013-3020, 2004.
181 Przulj N, Corneil DG and Jurisica I: Modeling interactome: scale-free or geometric? Bioinformatics *20(18)*: 3508-3515, 2004.
182 Przulj N, Wigle DA and Jurisica I: Functional topology in a network of protein interactions. Bioinformatics *20(3)*: 340-348, 2004.

Received November 28, 2005
Accepted December 19, 2005