Proteomics and Disease The Challenges for Technology and Discovery

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · APRIL 2004

Impact Factor: 4.25 · DOI: 10.1021/pr049958+ · Source: PubMed

2 AUTHORS, INCLUDING:

Thomas Fehniger
Lund University
109 PUBLICATIONS **2,056** CITATIONS

SEE PROFILE

Journal of proteome research

Proteomics and Disease—The Challenges for Technology and Discovery

György Marko-Varga* and Thomas E. Fehniger*

AstraZeneca R&D, Respiratory and Inflammation, Department of Biological Sciences, S-221 87 Lund, Sweden

Received February 9, 2004

1. Introduction

During the last five years, the term proteomics has found it's way into nearly all of the basic life science disciplines, including medicine. The popularity and interest level of studies promoting our understanding of the variety and forms of proteomes have fostered the introduction of several specialty journals to meet that demand, including this, ***The Journal of Proteome Research***. Our levels of understanding and our appreciation of the complexities surrounding protein expression, function, and detection have grown with a sophistication in applications of technology and a genuine understanding of the interactive networks of protein modules and pathways present at every level of biology.

If we take stock then in the lessons we have learned over these few years, we quickly realize that the supply and demand equation of current unmet need far outweighs our knowledge of biological protein systems. We are only beginning to learn our lessons.

But it is a truly worthwhile activity to review the measurements of our progress and to set goals which will have an impact on our thinking and the activities of our community.

What are some of the lessons we have learned? How will we use the knowledge we gain about single proteins or entire pathways to treat patients more effectively? How do we devise models to separate out functional differences in pathways shared between multiple cells, tissues, or organs? How do we sample proteins that are differentially regulated and expressed, in discongruent temporal phases? What are the "true" regulators of protein expression that initiate disease?

If the ultimate purpose of clinical proteomic studies is to relate protein expression profiles to specific disease phenotypes, it is important to understand the relationships between the biological process under investigation and the distribution of the components related to these processes. The overall dynamic range of individual protein levels varies over 10 orders of magnitude in the body.[1,2] In tissue, the rule is to try to discover relationships between cells and their products within functional compartments, and as close as possible to the foci of disease. Of course, not all diseases have foci or origins, which can be investigated by sampling. In the future, we predict that we will have maps that describe where one disease starts and where another disease ends. This futuristic guide will track multiple diseases and responses to treatment by measuring the constitutive elements of disease. But to begin, one must have a clear definition of the elements of disease. In our opinion, this currently poses one of the toughest challenges to research scientists studying Proteomics and Disease.

It is not easy to imagine the complexity of interactions, which involve all of the biological activities and processes which are studied using proteomics today. The main levels of study can be roughly divided into three components: expression, structure, and function. If we compose Venn diagrams to envision structure—function—expression activities segregated into schemes of health and disease (Figure 1) we find that we are required to include another ring representing subclinical or "symptom free" disease to account for those individuals which are neither healthy nor apparently diseased. This represents an important obstacle to overcome in any of our studies, that is, *How can we define the normal state in our surveys of protein expression? How do we differentiate the healthy phenotype from that of disease?* The wear and tear pathology of daily life must accommodate the fibrotic scarring from previous infections, the permanent deposits of carbon particles and soot within the lungs, as well as the changes in body mass index which occur with aging without invoking a disease definition. There are many factors and processes which define and qualify our definitions of health and disease such as the genetics and genomics of the individual, their developmental stage and capacity to respond to multiple stimuli from the external environment, the mechanical and structural limitations of cells, tissues, and organs for continual functional activity, their cellular capacity for activation, growth, repair, and regeneration using multitudes of inter-connected yet individually regulated processes, their unique immunological status and their susceptibility to biological challenge, and the degree of biological degeneracy and redundancy by independent elements for meeting the demand for functional connectivity. The sliding scale for calibrating the change from health to disease is not rigid. Rather, the scale is designed to allow a maximum freedom for individual variation and diversification within the sets of boundaries determined by these factors. The models we construct to encompass the transition between health and disease will likely contain additional elements that act at the local level and promote the balance between alternative phenotypic forms. If we just consider the expression component alone, we could conceive the natural divisions of high/low, local/distal, molecular/multicellular, loss/gain, episodic/continuous, and growth/atrophy. *How many different cause and effect or action and reaction relationships can occur in a disease?* The current solution for the maximum number of symmetrical

* To whom correspondence should be addressed. Telephone: +46 6 336000. E-mail: György.Marko-Varga@Astrazeneca.Com; E-mail: Thomas.Fehniger@Astrazeneca.Com.

10.1021/pr049958+ CCC: $27.50

Published on Web 03/18/2004

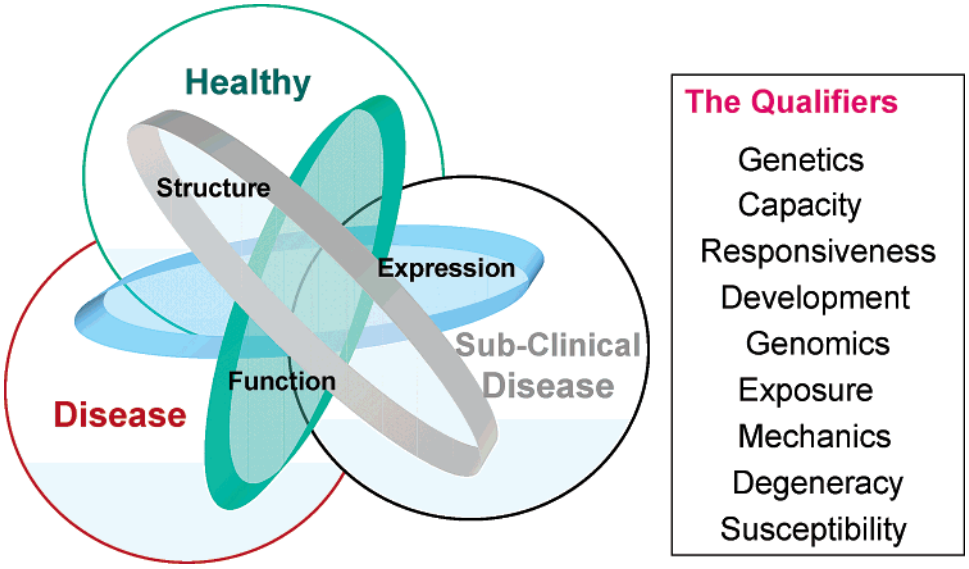

Figure 1. The Division of Health and Disease. The dynamic processes which define Protein Expression, Protein Structure, and Protein Function form the interactive environment which can be altered during disease. It is difficult to fully separate Health and Disease as unique divisions and Sub-Clinical disease functions as a gateway for either return to normality or continuation into disease. There are many factors which act to qualify the processes of structure, function and expression in each individual including genetic background, genomic organization, and variance, as well as the biological capacity for development and responsiveness to internal or external environmental challenge. Overall protein expression can be influenced by these and other criteria.

components within Venn projection is 11.[3] Perhaps there are also a maximal number of principle components within each disease we study.

This volume is dedicated to raise the level of awareness regarding current protein expression studies within diseases such as asthma, cancer, Alzheimer, cardiovascular and Chronic obstructive pulmonary disease (COPD), and inflammatory processes associated with many diseases. To achieve the level of understanding, which is required in such an undertaking, we are fortunate to have received the support of a unique set of contributions in this issue from many of the top strategists within the field. We are deeply grateful for their engagement and for giving us all an opportunity to share in their critical thinking.

The technology that supports proteomics is strong and comprehensive, and is matched equally against a community of talented scientists. We have required and attracted a broad spectrum of skilled people to power the development of Proteomics and Disease. While our ultimate tests will come in the clinic, the early development phase of clinical proteomics has relied heavily on technology, and especially on the expert contributions of engineers, physicists, chemists, bio-informaticians, and manufactures of instrumentation. We would like to begin by highlighting some of the important aspects of the special adaptations of technology that are used today to study disease.

2. Technology Platforms of Proteomics

Proteomics has been defined in many different ways. It is not easy to come up with a quick buzzword definition that adequately describes the detailed activities within the multitude of platforms which are being developed to tackle protein analysis today. Proteomics can broadly be defined as a collection of scientific approaches, and the application of technology toolboxes, to characterize the protein content within cell organelles, cells, tissues, and whole organisms. Interestingly, there was a recent re-phrasing of the proteomics definition, to include not only the expression of proteins as gene products, but also to include protein modifications as a natural mechanistic effect of cellular regulation.[4−6] The technology available for studying proteome expression and resolving exact protein and peptide identities in complex mixtures of biological samples allows global protein expression within cells, fluids, and tissue to be approached with confidence. This confidence is due in part to reproducible repetitive sampling and analysis technologies including robotics data acquisition and high level mass spectrometry including both laser-desorption and electro spray ionization. No doubt the development within the mass spectrometry field has been extremely important for the generation of accurate sequence and structural biological information. This was appropriately recognized recently with the Nobel Prize in Chemistry in 2002,[7] where both the electro spray, and the laser desorption ionization principles, the two most applied techniques within proteomics study today were recognized for their fundamental contributions in "*...the development of methods for identification and structure analyses of biological macromolecules.*"[8] An important goal of proteomics studies is to understand the biological roles of specific proteins and use this information to identify new drug targets. Proteomics research also has applications in target validation, drug screening, and the discovery of diagnostic markers, also presented in this volume by the research groups of Lefkovits, Hanash, Dunn, and Petricoin & Liotta.

The proteomics research areas within biomedical applications roughly segregate into four fields of categories of protein study: expression proteomics, functional proteomics, structural proteomics, and chemi-proteomics.

2.1. Expression Proteomics. Expression proteomics includes the traditional approaches for separating and quantifying the patterns of stochastic protein expression in samples including high resolution 2-DE gel analysis, multidimensional HPLC, capillary electrophoresis, and combinations of 1-DE gels with

nano-flow microcapillary liquid chromatography. Some of these approaches have been utilized for decades and have founded our basic understanding of the qualitative and quantitative relationships of protein expression levels. Typically, proteins are fractionated and then the isolated fractions are analyzed by MS, which assigns MS spectra identities to individual peaks. Using matched set analysis, we can compare the relative indices of abundance and calculate the degree of change in relative abundance for individual protein entities or groups of proteins. The Expression proteomics field has readily adapted new technology platforms to obtain better separation profiles and more exact identifications of the complex mixtures of proteins within biological samples. The recent development of isotope-coded affinity tag (ICAT) labeling which allows the simultaneous comparison of several samples together using multiplexing presents a new powerful addition to the toolbox. ICAT methodology is presented further in section 3.3 and in applications described by both Broder's group at Celera and by Hood's group at Systems Biology in this issue.

2.2. Functional Proteomics. Functional proteomics is the study of the function of proteins within biological systems and the regulation of their expression within these systems. Functional proteomics also includes studies of the affinity of small-molecule ligand-protein interactions, which are shared with the Structural Proteomics and Chemi-proteomics fields (see below). Functional proteomics addresses the activities within cells, which impact on the lifespan of proteins, such as their interaction with other proteins within these systems, their rates of expression and turnover, and the variances of form which impact on function. Following translation and synthesis, many of the proteins undergo post-translational modifications (PTM). The most common protein modifications occur by the addition of carbohydrates and/or phosphate groups, which affect both protein structure and function. These protein modifications play vital roles in modulating protein function which act to focus some activities while restricting others, as well as phosphorylation states which regulate and determine signal transduction and cellular activation.[9] The DNA does not code for the mechanistic aspects of protein modification per se. Rather, these modifications occur during the process of protein production, trafficking, and placement within the biological system. Interacting modification pathways (often enzymatic in function) drive PTM and can occur in either the intracellular or extra-cellular microenvironment: for example, within the extra-cellular spaces either attached to matrix structures or the surface of cells, or within the interstitial fluid, as seen in the processing of cytokines between latent and active forms. In essence, *all* modifications of a given protein occur within a biological context, and occur because of specific biological reasons, and that is why importantly we make specific annotations for each of the modified variants and functional forms of proteins.

Much of the current interest in functional proteomics is focused on developing our understanding of the relationship between the activities of specific proteins, their microenvironment, and their temporal expression during disease processes. The complex protein−protein interactions, which occur within pathways, often occur with protein complexes. A recent milestone in our understanding of the nature of protein−protein interactions was achieved by the analysis and mapping of the thousands of protein complexes present within yeast[10,11] This information will undoubtedly help us in devising strategies for similar mappings of pathways active in human disease.

Examples of functional proteomic studies are presented later in this volume by Ovaa and Ploegh and by Domon and Broder.

2.3. Structural Proteomics. Structural proteomics merges structural biology and protein chemistry with traditional cell based expression assays. Structural proteomics involves the determination and analysis of the three-dimensional structures of proteins which contribute to their function and their molecular identity to other proteins. Important information regarding locations of side groups, and the position and distribution of key domains can be obtained from X-ray crystallography and NMR studies. Today, the number of entries of structural solutions for proteins has increased every year and the Protein Data Bank (PDB) database contained 21 635 entries of peptides, proteins, and viruses and 983 entries of proteins/nucleic acids as of January 13, 2004.[12]

In combination with cell-based production using directed or positional cloning within specially derived cells in culture, we can further modify specific sites on the protein to study the contribution of selected structure/function relationships to activity. Thus, we can use the cellular environment to produce sufficient amounts of customized proteins for further study.[13] The information about the 3-dimensional structure of the protein is central to the design and development of small molecular weight compounds that bind to key functional groups. For example changes between active and inactive states, or vice versa.

The natural variation of protein sequence, structure, and function that occur in man now will be studied in detail to determine the genetic and inherited forms of protein expression. The completion of the human genome by the HUGO project in collaboration with the major partners: the National Human Genome Research Institute, the Wellcome Trust-Sanger Institute, as well as the sequencing centers at MIT's Whitehead Institute, Washington University in St. Louis, and Baylor College of Medicine has laid the groundwork for studying the variation in phenotype and genotype of the proteins identifies in disease.

Structural proteomics approaches are being used to discover novel drug targets. The direction the area is taking commercially is toward high-throughput structural proteomics methods for structure determination. These structures could then be matched against complementary mini-libraries of compound classes matching the targets on the protein with small molecular weight docking partners. This approach of structure based drug design is in it's early phases,and will benefit by improved capacity in the delivery of structure-drug pairs for study and in the establishment of proof of drug-target affinity mechanism studies in experimental settings.

2.4. Chemi-Proteomics. The area of Chemi-Proteomics picks out the identities of proteins that interact with small pharmacologically active compounds under defined conditions. Similar to affinity chromatography where immobilized ligands interact with functional groups on proteins within complex mixtures, chemi-proteomics uses small molecules to identify complementary "targets" bound at the active site of the proposed biology. This is a newly defined area where drug-target interactions have been the principle components of study.[14] In a sense, the technology can be compared to photoaffinity labeling and classical organic synthesis of isotope labels into drug molecules revealing sites of near contact and binding to the target molecule(s). This method is being followed carefully in studies validating the technology in applications of the following: (i) known target-drug interactions, or, unknown (black box) molecular interactions with drug candidates; (ii) The capacity to be selec-

tive for binding the correct protein in terms of affinity and definitions of specificity, and (iii) The appropriateness of the protein library being surveyed in terms of the biological coupling to disease mechanisms (Proof of Mechanism) or the linkage association with actual disease processes (Proof of Principle).

To a large extent, the chemi-proteomics initiative relies on the availability of small molecule libraries which in theory would need to cover the entire spectrum of three-dimensional space filling in order to meet all possible candidate interactions with potential docking sites on proteins and their glycosylated, phosphorylated, ubiquitinated, nitrated, or sulfated substituents. It takes a repertoire of some large size to meet this demand, but there is precedent in the repertoire of antibodies. In man the diversity of antibody binding sites is calculated to be 10.[22] The number of unique chemical entities in the small molecule libraries which exist today represent only a minute fraction of this coverage.

3. Technology for Analyzing Protein Expression

3.1. Quantitative Protein Expression Analysis. The experimental design of studies comparing health and disease at the molecular level, often rely upon global surveys of gene and/or protein expression. The discordance between gene expression and protein expression is well established.[15−17] These studies have compared mRNA versus protein expression levels in a variety of contexts but have failed to show statistical correlations. We then are left with the complicated task of relating levels of protein abundance to fundamental functional biology.[18,19] This is a real challenge since (up to 90%) of proteins are believed to be present as low copy numbers.[18] We can make a simple calculation to estimate at what relative level of abundance we are obtaining protein expression data today. First, we know that not all proteins identified as being present at a given time will be expressed as identical unit forms, or in identical functional forms. Second, the number of human proteins as gene-products can totally range between 300 000−3 000 000 species, depending upon splice variants and the numbers of multiple proteins encoded in single genes.[18] What we need to encompass, within an abundance equation, is the post-translational modifications occurring within the cell that would generate a Cell-Protein-Index Number (CPIN) which is very difficult to estimate. For instance, if one considers that there are 30-types of phosphorylation variants of a single phospho-protein (∼1200 kinases have been identified!!), as well as a 100-fold forms of glycolsylation possibilities of a single glycoprotein, the CPIN could vary from a few million to several hundreds of million different protein forms within a cell. Third, studies calculating the dynamic range of protein expression within cells and/or plasma estimate that the differences between the levels of most and least abundant proteins are on the order of 10^{8-10} in magnitude.[1,2] If we would assume then that in typical expression analysis studies, total cell input numbers are present at most, not more than tens of milligrams of starting material, then the least abundant proteins would be present at starting levels not exceeding picogram levels. If we put the equation together, then we are left with the ***fraction of relative abundance*** within each expression profile

$$\frac{\text{input cell number/protein concentration} \times \text{CPIN}}{\text{dynamic range}} = \text{expression profile}$$

However, the proteins within this fraction will change with every cell type analyzed and every biological context investigated over time. Considering our analysis performance using gel technology and/or liquid phase separation technologies which typically provide annotation matches in the order of thousands of identities (out of estimated total $>10^{6-8}$), we as yet, only map a small fraction of the entire biological activity within cells. Even so, in our experience, a fractional identification of the proteome is valuable in assigning contextual relationships between protein patterns derived from tissue and a specific biological response. For example, using an experimental allergen challenge model, it was possible to notate within whole tissue, an "allergic fingerprint" of protein expression at medium-, and lower abundance levels, despite the relatively low (300 μg) total amount of protein used in the analysis. (Presented later in this volume, Fehniger et al.) We believe this is due to the relative increases of proteins accumulating during the inflammatory process within tissue micro-environments, and hence at higher than normal local concentrations. The take home message here is that all biological processes will likely produce their own protein fingerprint at the local sites of disease. The further away one gets from this cell compartment, the more the fingerprint will be changed by dilution and addition from other sources contributing proteins into the sample module.

3.2. 2D-Gel Separations. 2-D PAGE as a protein expression separation media is the most common tool used today to characterize and proteins within biological samples. 2-D PAGE offers high resolving power, economy, and wealth of experience from previous study. 2D gel analysis coupled to MS identification processes have powerful resolving capacity: (i) in the order of 3−8 thousand annotated protein spots for metabolic labeled studies utilizing both S-^{35} and P-^{33} isotopes[20−22]; (ii) large sized 2D-gels, where the surface area of the separation matrix provides a resolving power up to 10.000 protein spot annotations. (iii) Utilizing narrow PI-range units (also named as Zoom gels) where the loading capacity of these systems allows milligram amounts of total protein to be run. This results in an improved annotation power in the lower abundance levels annotations. It is possible to run 2D-gel expression maps in a turn around cycle of 20 h, revealing a given characteristic fingerprint for any given cell model or human clinical material, certifying the expected characteristic protein map. The method provides intact isomeric forms of protein that are useful in studies detailing PTM. This is an advantage that 2DE holds over many other approaches. The small structural differences resulting from PTM are typically due to chemical variations of: phosphorylation, sulfatation, sulfation, glycosylation, nitration, and deamidation to mention the most common modifications.[9,23] Although chromatography based separation methods are gaining increased use and popularity in biological research, 2D-gels will continue to be used by many laboratories in the world.

There is a major challenge when it comes to the identifications of proteins analyzed by 2D-gels, and that is the digestion and extraction of the resulting peptide fingerprints from assigned protein gel spots. It is simply the limitation in the Michaelis−Menten enzyme kinetics that makes low level substrate (protein spots) concentrations extremely difficult to be cleaved by tryptic digestion. Currently, the only way to get around that problem is to increase the amount of biological sample applied to the assay, and preferable use narrow-range PI-strips to allow better resolution in a small PI-window. This ultimately will increase the resolving power and protein concentration of unique proteins in each single spot.

3.3. Multi Dimensional Chromatography Separations. Liquid Phase separations are increasingly utilized in protein

expression studies using combinations of multidimensional separation mechanisms. These tandem separation steps allow several physical-chemical properties to be selectively targeted. This results in achieving sharp separation and concentration gradients of both proteins and peptides. The addition of robotic sample delivery, solvent dispersal, sample collection, and semiautomatic MS sample analysis as modular platforms interfaced to each other allows real-time and fully on-line separations.[24−26] Using this approach, we have recently identified several thousands of peptides.[26]

The qualitative identity of proteins is achieved by mass spectrometry, where various ionization principles are applied, such as matrix assisted laser adsorption ionization MALDI, as well as electrospray ionization (ESI).[27−29] Fourier transform ion cyclotron resonance (FTICR) mass spectrophotometry instrumentation currently has the highest resolving power and is also starting to make an impact on the proteomics field. Although initial studies utilizing this technology data look promising, the final assessment will need to be proven in validated comparative studies.[30,31]

The complementary nature that ESI and MALDI offers is attractive. With ESI it seems that the peptide size is directly linked to the sequence ID, whereas MALDI TOF−TOF instruments seem to have improved performance with larger peptides.[32]

Samples applied to the MS are identified by peak distributions and atomic mass. Each peak identified above threshold detection is then assigned an atomic mass identity which can be used as the signature for the exact amino acid residues present in the protein/peptide fragment. Peptide peak mass identities are converted into protein identities by querying various database sources that assign the correct sequences to the MS-spectrum data.[4,6]

The advantage of liquid phase chromatographic separation systems (LC) is that the automated operation is much easier to perform. Compared to 2-DE gel analysis, which typically requires experienced hands to handle raw forms of the sample in multiple steps and with careful attention to detail, high order liquid phase separations are simpler and can be achieved even by less experienced operators. In the 2-DE analysis process it is relatively easy to introduce variegation effects because of technical or operator related causes, whereas in a modular LC process the sample is sealed until it leaves the system in collection fractions.

A lot of excitement has also been generated by the "shotgun proteomics" first described by Yates' group,[34,35] in which mixtures of proteins are proteolytically digested before separation using multidimensional LC in combination with MS and/or tandem MS/MS. In its common application, digests of total protein, which contain stoichiometric amounts of peptide constituents of native protein forms are applied to the LC system. One reason that LC systems, are successful using this approach is due to the simple fact that the Michaelis−Menten kinetics are favorable: that is to say, the substrate concentration is high, the digestion occurs in solution and at optimal temperature and pH, and the total amount of protein which is digested favors reactions with certain substrates. Increasing the levels of peptides available for analyses ultimately results in an improved identity annotation in protein expression studies.[36−38] The practical utility of using multidimensional column chromatography systems will be forthcoming in studies of model systems, such as with the yeast *Saccharomyces cerevisiae*. To date, of 6000 potential ORF encoded proteins in the genome, approximately 1500 have been identified using LC-MS approaches. The theoretical yield of over 250 000 total peptides will press the technology to deliver a comprehensive identity map[10,39−41] The fine-tuning of multidimensional chromatographic separation mechanisms will need to be harmonized and interfaced optimally to the mass spectrometer as the sequencing unit of the platform. There is an intrinsic complexity in the peptide components contained in each fraction analyzed by MS (Figure 2). There is a price to pay for running profiling studies under tight time constrains, and the price is lower annotation numbers and incomplete resolution: both in terms of separation, peak sharpness, and mass-, sequence identities. The other alternative is to devote the time required in order to generate a large fraction set of a given sample and then process all hundreds/thousands of fractions through the LC−MS platform. High annotations are thereby achieved, however the price to pay this time is the poor sample throughput. Probably, one way forward is the miniaturization of processes where micro/nano-separation microfluidic platforms are applied with parallel processing.

Quantitative measurements of protein expression are also favored in multidimensional liquid-phase separations through the application of isotope labeling.[42] This powerful development allows back to back comparisons for example between two individual sample types,such as pairs of nonactivated/activated, healthy/diseased, or treated/untreated biological samples.[43] The leading edge today is with isotope coded affinity tag (ICAT) labeling using either a heavy or light reagents to differentially distinguish pan-protein expression profiles which are run in tandem.[44−46] Differential quantization utilizing isotopes is enabled by the chemical labeling of all protein species through derivatization, metabolic labeling, or by enzymatic labeling.[44] For ICAT labeling, the LC separated differentially labeled peptide pairs from each digested peptide are quantified measuring the isotope ratios by mass spectrometry using stable isotopes as internal identity indices.

In retrospect, the development of high resolution capillary separations and proof of principles in complex sample analysis were accomplished by Csaba Horvath's group at Yale already in the early 70s.[47] Horvath was one of the pioneers in developing high performance liquid chromatography "HPLC" and already then performed high-resolution peptide-, and protein separations in capillaries. These separations were performed at high temperatures with impressive speed and resolution,[47,48] as well as at low temperatures for using capillary electrophoresis.[49] When asked about the dangers of denaturation and polymerization during >200°C separation temperatures Horvath retorted empirically, "They have no time! Do the calculations. They have no time!" (Gyrgy Marko-Varga, personal communication and ref 48).

There is no doubt that miniaturization will be a major issue, and matter of attention, in driving toward the goal of reaching lower threshold levels of protein expression analysis in biological studies.[50,51] This can be demonstrated in recent developments utilizing nanoliter flow rates in microcapillary chromatography column platforms. Most recent, are the developments in the use of monolith columns with 100 μm ID sizes, which negate the need for column frits.[52,53] The developments of silicon microchip and micro-fluid configurations are new approaches for establishing ways of making the separation more efficient.[50,51,54,55] Recently, Baba et al. presented fractionations from Jurkat cells using polymeric microchips performing the separation in nondenaturing conditions.[56] The only way to

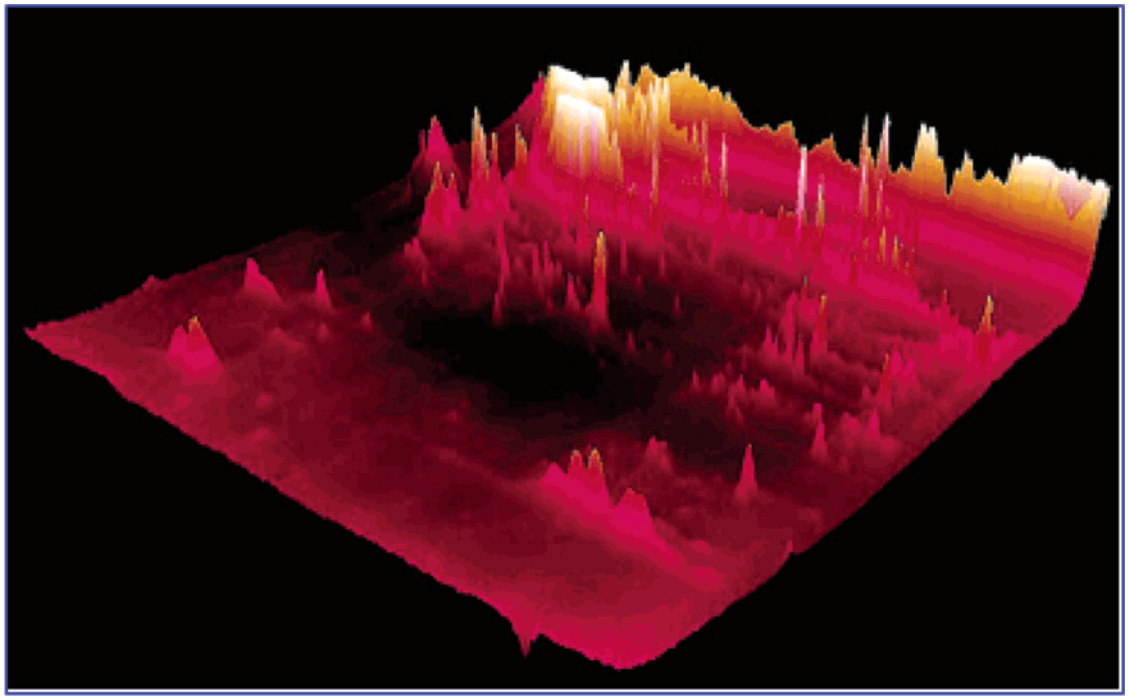

Figure 2. Pushing the Limits for Detection. The components of singular fractions separated by 2-DE or liquid chromotagraphy systems often are complex mixtures of many proteins and peptides. We will need to address new methods which will allow high resolution separations, coupled with speed, and accuracy in order to fully identify the species present in low abundance.

handle minute amounts of sample is by decreasing the sizes of the analytical scale whereby the samples are handled, separated and processed. Micro-sized technology is something that is available to some extent from various vendors, and (as proposed by Weston and Hood in this issue) on lab-on-a-chip micro-conduit developments. The next step in development is to take it to a nanotechnology scale, but here we still need to pass a build-up phase first, although, some scientists run nano-flow liquid chromatography and dename it as nanotechnology applications!

3.4. Protein Micro Arrays. The Protein-Chip Array development activities are taking similar routes into laboratory application as the DNA-chip array in the past. The difference is that protein microarray assays allows simultaneous identification and quantification of a large number of proteins from a small volume of any given sample within a single run, with the actual biological affinity/activity read-out.

Currently, there do not seem to be chip arrays that allow the prediction of high affinity protein capture molecules only on the basis of their primary amino acid sequence.[57] Lacking array products that arises from high density protein arrays, it seems that the application of phage display libraries, or other types of antibody libraries is not a trivial step to make: that is getting all the conditions equal for the antibody/antigen immuno-complex assay to "dance at the same time". Still, there are protein arrays that seem to be analytically useful in disease settings as discussed recently in studies using differing approaches and readouts.[59−61] The work of the Hanash group presented later in this volume provides a good example of the current state of the art using these approaches.

4. Relating Protein Expression to Disease

4.1. Linking The Technology Toolbox To The Clinic. The toolbox of technologies that we have outlined above, empower and provide us with a variety of approaches for studying the molecular and biological bases of disease. Today, we can study the dynamics of protein expression within a variety of biological compartments ranging from sub-cellular sites to the whole organism. By understanding key pathways functioning in the maintenance of steady-state function, activation, proliferation, apoptosis, repair, growth, and regulated gene expression, we hope to discover the fundamental roots of disease, and the possible sites for medical intervention strategies. The assignment of protein identities which are linked to key biological mechanisms associated with disease processes and disease progression are an important area of activity.

The unmet challenges of proteomics activities today revolve around understanding the roles which individual proteins or entire pathways play in the initiation and development of complex diseases,[62] or for that matter even simple diseases.

What are the process steps we are taking to meet these challenges? We believe there are five basic points which need to be attained: (1) A betterment in our general understanding of the pathophysiological basis of disease. (2) A linkage between clinical investigations and basic science research which provides quality clinical samples for study and standardized clinical measurements for correlation to biological activity. (3) Common searchable Databases of annotation for both peptides and proteins which are cross referenced to biological processes occurring in human disease and linked to data acquired using technology platforms traditionally not included in "protein only" databases. (4) A linkage between diseases which allows large datasets to be compared to common reference proteomes, be it cell or organ. This would include the development of statistical tools and approaches for comparing large datasets in unbiased testing. (5) The development of technology which will allow us to break the "abundance barrier".

The first two points are linked with a common purpose to raise the standard of our activities to their highest level. For

studies requiring human clinical samples, ***Best Practice***, must be our community's standard procedure for attaining clinical samples for our studies, starting with informed consent by the individual donor and with full ethical approval from the appropriate institutional review boards.

In order for us to segment health and disease, we will require a necessary and earnest contribution from the clinic, our clinical colleagues, and the clinicians among us, to set the stage for study with information about disease, about patients, and about the measurements used to assess their progression and evaluate their responses to treatment. The focus of clinical proteomics studies today is to discover direct linkage between protein expression and function and specific contexts of disease. The patients entering the clinic often can be diagnosed by pathologies associated by declines in function, such as the emphysemic lesions in the lungs of a COPD patient or the joint destruction associated with progressive arthritis (Figure 3). *Are there meaningful corollaries between specific patterns of protein expression and predictable clinical phenotypes?* We have good historical examples relating the measurements of singular proteins in the context of specific diseases such as Bence−Jones immunoglobulin light chains in multiple myeloma, or the deficiency of α-1-anti-trypsin activity with the development of emphysema. In the past few years, the studies of complex diseases have provided a different view of protein expression phenotypes in disease, in which multiple proteins and pathways are related to structure−function−expression levels and activities. In Table 1 are listed representative examples of proteomes associated with specific clinical phenotypes. Although we are at the beginning of fully relating proteome character and disease phenotype, we have already made some remarkable associations in many of the major diseases including cancer, heart disease, neurodegerative diseases such as Alzheimers, and Huntingtons Disease, arthritic diseases and in several cases where histological classification was unable to resolve disease identity or stage. Studies have already begun to identify the protein phenotypes which are associated with disease progression and outcome including tumor invasiveness, metastatic potential, and long term survival. We have also begun to relate clinical proteomic phenotype to protein expression to clinical measurements of disease by imaging as shown in the studies staging articular erosion during rheumatoid and osteoarthritis. The clinical information delivered with such methods as CT, HRCT, MRI, PET, and γ-scintography can be used to noninvasively measure structural and functional changes at the very foci of disease. Clinical measurements such as these imaging technologies, provide additional context to disease progression and are important landmarks for comparing patterns of protein expression occurring focally within tissue.

Pathology and Histology as disciplines, and Pathologists and Histologists as investigators, have a central role to play in developing deeper insights into the biological processes of disease within the compartments of cells, tissue, and organs which we study. It is here that we will develop answers to the questions of "*What to study? Where to study? And When to study?* Of course, these questions have been posed long ago. We would like to think that Lewis Thomas, a pathologist, would feel very much at home today within the Science of our proteomics community.[63] We think he would surely support the efforts with HUGO and HUPO to link gene and protein expression to disease. We can only imagine his delight in the results of studies mapping protein pathways and networks within lower organisms, such as the large scale analyses of genes to proteins in *Caenorhabditis elegans*,[64,65] the studies to characterize the multi-protein complexes in *Saccharomyces cerevisiae*,[10,66] the resolution of life stage proteomes of the Malaria parasite[67] (and see Yates this issue), and the recent protein interaction map of *Drosophila melanogaster*.[68] The main customer of the data generated in these and like studies will be those of us who have their eyes, like he did, on human biology.

The third and fourth steps to meet the challenge are linked to the unmet need of linking protein expression identified by MS to the context of biology which is being investigated. Experimental models of disease have made a significant contribution to our knowledge of the regulation of protein expression during induced, early onset, or late onset diseases (Figure 3b). In addition genetic models such as knock outs, knock-ins, and transgenic animals have also been invaluable in determining the contribution of individual proteins within biological processes. The Jackson Laboratories have pioneered the establishment of databases correlating genetic mutations and disease phenotype in strains of mutant mice as well as their pathological characterization. The establishment of such key reference proteome systems representing the dynamic changes in protein expression during disease will be vital to the interpretation of changes observed in specific samplings of disease states and specific cells obtained from these samples. The creation of reference databases of proteins will allow us to assign context to the protein sets identified. An example of context: the identity returns from the database as "epithelial associated protein". But which type of epithelial cell express the protein?, Which tissue and which organs ? What stage of differentiation or activation? Are they normal or neoplastic? What other types of cells express "epithelial associated proteins". The "smartness" of the algorithms available for mining large data sets of peptide masses, in an automated format, is under constant improvement and validation. The success rate, whereby proteins are identified correctly and in context to the biological system from which they originate, is highly dependent upon the annotation records associated with the proteins present in the database. Since proteins often are identified by novel naming conventions assigned to unique sequences but expressed by proteins derived from several cell or even species types, it is common to find several annotation identities assigned by the database engines to individual MS peaks. Choosing the correct protein identity from these annotations thus is a formidable challenge.

Systematic approaches which capture and disseminate proteomics experimental data by "reading biology" are the ultimate ambition for unraveling the complexity of biological systems at hand. There are proteomics experimental data repository tools allowing sample processing, capturing and structuring expression annotation and identity (www.mged.org/Workgroups/MAGE/mage.html) as well as other software packages such as (www.expasy.org). New approaches of assignments of protein expression to functional classes on the basis of their network of physical interactions was recently presented using two hybrid experiments.[69] There are also organized efforts to develop annotation and terminology for classifying and assigning proteins into groupings of common protein families established using ontological association scoring from the Gene Ontology Consortium.[70] Many of the questions and challenges associated with the classification and mining of proteins within large datasets are considered by Weston and Hood, in this issue.

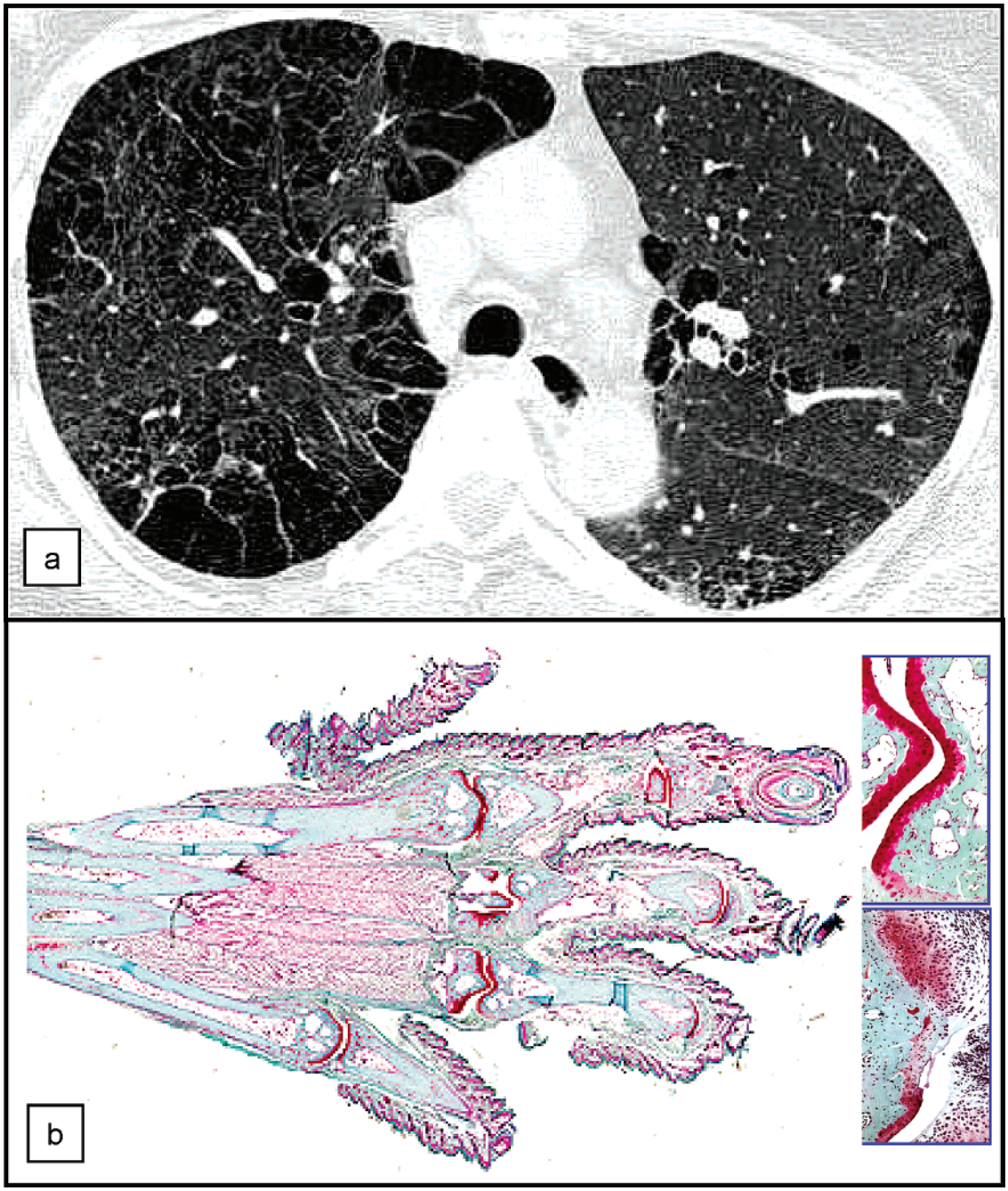

Figure 3. Clinical Measurements Provide Disease Context to Proteomics Study. To understand the changes in biological function which are active during disease, we must rely on clinical information which reflects altered functional states. Medical imaging is a powerful tool for studying the structural changes which occur locally during disease progression. In (a), the image is a high resolution computer tomography (HRCT) image of a cross section of the chest cavity which detects paraceptal and centrilobular emphysema within the lobes of the lungs of a COPD patient. (Image courtesy of Dr. Sahir Shaker and Prof. Asker Dirksen, Gentofte Hospital, Copenhagen.) This information provides valuable context to studies of protein expression during ongoing disease. Experimental Disease Models also provide valuable context for interpreting changes in structure and function which occur during progressive disease. In (b) is shown a cross sectional image of a normal mouse paw revealing, muscles, bone, and joints. The cartilage seen staining crimson within the joints undergoes a destructive inflammatory process during experimental arthritis caused by the induction of proteinases by the activated infiltrating leukocytes. Clinical models play an essential role in developing our understanding of regulated protein expression during health and disease. (Histological preparation by Catharina Zackrisson, AstraZeneca R&D.)

The fifth step needed for attaining global profiling of regulated protein sets is to "break the abundance barrier". How can we measure trace amounts of proteins in complex mixtures? This problem has been a matter of discussion for some years now and is approached seriously by a number of the contributors to this issue. We expect that breakthroughs are forthcoming but leave it as a point for us all to consider.

The sequencing and structural analysis of entire, or parts of, proteomes will help reveal cellular functions of proteins as well as how they assemble into molecular machines and networks. It is clear that human proteome studies have opened up the biological understanding of molecular and cellular processes for example receptor signaling pathways[71−73] as well as the human spliceosome, a multi-protein complex regulating

Table 1. Proteome Expression and Clinical Phenotype

proteome	clinical phenotype	literature reference
cardiac muscle	dilated cardiomyopathy heart failure	Pleissner, K. P. *Electrophoresis* **1997**, *18*, 802; McGregor, E. *Human Mol. Gen*. **2003**, *12*, R135; Dos Remedios, C. G. *J. Muscle Res. Cell. Motil.* **2003**, *24*, 251; Van Eyk, J. *Curr Opin Mol Ther*. **2001**, *6*, 546.
lung tumor tissue	protein profile versus histology	Yanagisawa, K. *Lancet* **2003**, *362*, 433; Yanagisawa, K. *Clin. Lung Cancer* **2003**, *2*, 113.
lung normal tissue	primary versus metastatic tumor nodal involvement clinical outcome/survival	
lung tumor tissue adenocarcinoma	clinical outcome/survival	Chen, G. *Proc. Acad. Natl. Sci. U.S.A.* **2003**, *100*, 13 537.
ovarian tumor tissue	invasive clinical phenotype	Jones, M. B. *Proteomics* **2002**, *2*, 76.
ovarian tumor serum	ovarian cancer identification	Petricoin, E. F. *Lancet* **2002**, *359*, 572; Zhu, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14 666.
colorectal epithelium	colorectal cancer identification versus normal	Ott, V. *Pharmacogenomics J.* **2001**, *1*, 142.
prostrate epithelium **prostrate stroma**	tumor phenotype/progression	Grubb, R. L. *Proteomics* **2003**, *11*, 2142; Cazares, L. H. *Clin. Cancer Res.* **2002**, *8*, 2541; Zheng, Y. *Cancer* **2003**, *98*, 2576.
liver tumor tissue	hepatocarcinoma identification	Kim, J. *Electrophoresis* **2002**, *23*, 4142.
liver normal tissue		
liver disease serum	differentiating hepatocarcinoma from chronic liver disease	Poon, T. C. *Clin. Chem*. **2003**, *49*, 752.
brain normal tissue **brain tumor tissue**	differentiation of glioma by histology grade	Schwartz, S. A. *Clin. Cancer Res.* **2004**, *3*, 981.
brain tumor tissue	differentiation of histologically identical hemangioblastoma from renal clear cell carcinoma	Vortmeyer, A. O. *Neurology* **2003**, *61*, 1626.
brain tissue amyloid-β protein τ protein α-synuclein	neurodegenerative disease/dementia	Cummings, J. L. *Ann. Neurol*. **2003**, *2*, 147; Rohlff, C. *Curr. Opin. Mol. Ther.* **2002**, *4*, 251; Butterfield, D. A. *J. Neurochem*. **2003**, *86*, 1313.
cerebrospinal fluid	differentiating viral and bacterial meningitis	Glimaker, M. *Scand. J. Infect. Dis. Suppl.* **1992**, *85*, 1.
rheumatoid arthritis **synovial fluid**	clinical staging/early articular erosion. metalloproteinase levels versus radiographic imaging	Cunnanne, G. *Arthritis Rheum*. **2001**, *44*, 2263; Ishiguro, N. *Arthritis Rheum*. **2001**, *44*, 2503.
osteoarthritis **synovial fluid**	clinical staging articular erosion. metalloproteinase levels versus MRI imaging	Ishiguro, N. *Arthritis Rheum*. **1999**, *42*, 129.
synovial fluid	differentiating osteo- and rhematoid arthritis	Sinz, A. *Electrophoresis* **2002**, *23*, 3445; Uchida, T. *J. Proteome Res.* **2002**, *1*,495.
urine	renal allograft rejection	Clarke, W. *Ann. Surg.* **2003**, *5*, 660.

mRNA processing.[74] Signaling kinetics is an important variable in pathway biology that is affected by sequential multisite phosphorylation that may give arise to switch like activities for kinases[73] Many of these in vitro model studies have resulted in the generation of large biological databases that are highly useful to build hypotheses in the understanding of disease evolvements. For example, the EGF-receptor kinase pathway has been studied in quite some detail to describe the numerous protein−protein interactions occurring within the pathway process of receptor mediated signaling leading to cellular activation.[75,76] Further activities which have led to improvement in annotating protein−protein interactions, have also been valuable such as the studies to optimize the identification of weak functionally important binders in pathways, while controlling background noise.[76]

Protein Expression analysis of tissues has become a standard method to highlight quantitative and qualitative differences between sample sets representing different phases and stages of pathology. Today Laser Capture Microscopy technologies have revolutionized our ability to isolate relevant cells and tissue components for study.[77−80]

Typically, tissue is heterogeneous in its molecular and cellular composition with respect to both macro-(orientation of the sample within a larger tissue or organ), as well as microenvironment (the breadth of cellular structures, organized within identifiable compartments related to function). In another part of this volume (see Fehniger et al. later in this volume) we have discussed in detail many of the aspects concerned with sampling of tissue versus experimental design. By combining LCM with detailed histological characterization we can assign true biological context to the quantitative and qualitative aspects of protein expression. An alternative approach to LCM, known as *tissue imaging*, has also provided key data for understanding the content and context of local protein expression within clinical biopsy material. This method holds a niche not approached with other technologies, not the least of which is the ability to measure proteins, peptides, and drugs/compounds simultaneously. (See separate articles by Caprioli and Andren later in this issue)

4.2. Biomarkers. Many of the modern approaches for studying disease compare steady-state functions such as repair, growth, and regulated gene expression within the various

biological compartments organized by specialized function, be it mitochondria or blood vessels. The assignment of protein identities which are linked to key biological mechanisms which are associated with disease processes and disease progression are an important area of this work. One of the principle areas of development with the community today concerns the identification and validation of biomarkers (mechanistic, disease, surrogate, toxicology) which can be used as measurements within clinical studies. The further application for predictive diagnostics is another hot area where proteomics is moving forward. This area is considered by many scientists and physicians as a holy grail to for characterizing human diseases and disease states. Biomarkers are here to stay.[77,81−83]

Recently, the use of biomarkers within clinical settings in order to support claims within Phase 2 (proof of principle) as well as Phase 3 (proof of concept) drug studies has gained increased importance.[81,82] The collaborative efforts by the FDA and the NIH in a joint biomarker screening study on human blood samples from cancer patients illustrates an example where the research team could verify fragmentation similarities differing significantly between healthy and the cancer patients.[83] This report heralds a new approach for defining biomarkers in that the mass identity(s), even of unknown protein or peptide annotation identity(s), can be used as a measure for monitoring disease. The high accuracy and precision that this technology offers provides sufficient argument for further studies validating the approach in semi-diagnostic settings.

There have been a number of papers recently in which biomarker candidates have been identified. (See manuscripts by Petricoin and Liotta, Hanash, Lefkovits, and Dunn.) This is an extremely intensive research area at the moment where both the FDA as well as the European drug administration IAEA are closely following the developments within the area. It is envisioned that the outcome of these research activities will have a strong impact on modern drug development. Key questions relating to the use of proteomics in the staging of disease, or the association of specific protein expression patterns with clinical outcome are widely being discussed. If predictive indices can be attained then this will have wide ranging impact on drug development and ultimately to the improvement of quality of life for large patient groups. NIH is also moving forward on the recent developments, unravelling the "road map for medical research" program with a comprehensive set of new strategies, in which the agency will pursue basic and clinical research. Free access to data generated in these biomarker activities generated from clinical studies is also a remit of the new NIH policy.

Accordingly, FDA recently challenged a large group of pharma-, and governemental representatives at a global meeting, discussing the value of pharmacogenomic data in regulatory decision making.[83−85] A guidance paper regarding pharmacogenomics has been forwarded by the FDA to industry that outlines a set of nonbinding recommendations. It is expected that the various biomarker activities within pharma/biotech/clinics will be expanded remarkably to reach an estimated value of $ 9 billion by 2008.[85] There are currently several feasibility studies underway to test the utility of MS based clinical assays for performing biomarker discovery studies. These protein profiles will help determine which patients are likely to respond to specific drug treatments, in order to optimize the likelihood of successful treatment outcome. The major gain here will be to match directly appropriate therapy to predicted response from the onset of drug treatment. This will have significant impact on dosing and treatment regiments and reduce the possibility of noneffect outcomes which is not uncommon in everyday medical treatments today. (See Petricoin and Liotta presented further in this issue for a detailed discussion)

The relationship between protein expression, regulation and function within diseases has recently generated a global activity, the HUPO initiative (www.HUPO.org) that includes five platforms including plasma, brain, liver, bioinformatics and instrumentation. Recent activities also address global protein function predictions, with the analysis of protein−protein interactions, being one of the most challenging tasks within the post-genomic era. (See Hanash later in this issue for a detailed discussion of the current HUPO inititive). The task is a real challenge since the number of human genes and multi coding regions in genes will result in hundreds of thousands of proteins. By adding the PTM effect, the number of human proteins will reach millions of individual structures. Additionally, the HUPO plasma project initiative is currently involved in defining the plasma proteome in a "round robin" exchange of standardized samples between 50 consortium centers spread out on all continents.

A final word on biomarkers. Gerald Edelman has described biological "degeneracy" as the ability of elements that are structurally different to perform the same function.[87] Degeneracy is a prominent functional property in many biological systems (examples are receptors with more than one ligand, ligands with more than one receptor, etc) In our hunt for biomarkers, we have singularly focused on individual molecules as the read outs of activity and function. But is this founded from biological principle or from reductionist thinking? Degeneracy is the hound that stalks behind our hunt for biomarkers. The alternative of describing singular markers is to claim entire modules of regulated proteins as our biomarkers. We recently reported a study in which proteins present in the bronchial lavage of healthy never smokers and smokers could be segregated into respective groups by principle component analysis using all 944 proteins identified in the study.[88] We believe that analyses of entire protein modules detected in clinical samples will allow unbiased association of abundance levels with disease.

5. Future Perspectives

It is envisaged that direct links between pathophysiology state and protein expression state will shortly be attained for a number of important diseases. The study of tissues at the single cell level and in context to the microenvironments of disease will help us to decipher the complexity of the biological readout we make today. Complementary protein expression profiling of biofluids sampled from several compartments will help shed light on the evolvement of local and systemic diseases. This is of great importance since it will make it possible to scale the levels of disease and allow us to establish characteristic fingerprints of the modules of protein expression associated with disease state. The understanding of the cellular and molecular operators separating mild-moderate and severe-chronic disease are pathological challenges that still remains to be solved for many diseases. In vivo models play a central role for unraveling disease for it is here that we can measure and test hypotheses associating protein-target-function and drug intervention. It is also envisioned by us that the technology development will drive the biology forward making the experimental exercise of mapping the proteome of human

disease possible. In this context, glycomics, or carbohydrate chemistry applied to understand glyco-protein modifications, and the alteration between active and inactive affinities to bio-adsorptive cellular functions has been addressed by novel approaches.[89,90]

A major challenge ahead still remains the access and use of the bioinformatic output of proteomics studies and the understanding of the intrinsic biological meanings within that data. This challenge has also been described as the "Achilles' heel of proteomics" by Patterson.[91] Surely, the wide adoption of proteomics in life science will require developments that can combat data overload and ensure the quality of the data generated. Prediction of human protein function according to gene ontology categories is one approach for treating and restoring Achilles[92] The International Human Genome Consortium announced late April last year that the Human Genome Project, in which the 3 billion base pairs of the human genome were sequenced by a team of researchers spread across the globe, is complete. This milestone will act as an even stronger driver of the proteomics research area in the future.

Niels K. Jerne once wrote a paper entitled "The Complete Solution to Immunology".[93] Many years late, Immunology as a discipline has become more or less integrated into most branches of cell biology and is no longer seen as a stand alone science paralleling alchemy in it's experimental mixtures of cells and species. Today we all agree that Immunology is pervasive. We hope that Proteomics will enjoy the same enthusiasm and acceptance and provide an equal share in contribution to our ability to diagnose, follow, and manage disease.

References

(1) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics.* **2003**, *2*(1), 50.
(2) Kettman, J. R.; Coleclough, C.; Frey, J. R.; Lefkovits, I. *Proteomics* **2002**, *2*(6), 624−631.
(3) Cipra, B. *Science* **2003**, *299*, 651.
(4) Aebersold, R.; Mann, M. *Nature* **2003**, *2*, 198−207.
(5) Phizicky, E.; Bastiaens, P. I.; Zhu, H.; Snyde, M.; Fields, S. *Nature* **2003**, *422*, 208−215.
(6) Kenyon, G. L.; DeMarini, D. M.; Fuchs, E. et al. *Mol. Cell. Proteomics* **2002**, *1*, 763−780.
(7) The Nobel Foundation http://www.nobel.se/chemistry/laureates/2002/index.html.
(8) Fenn, J.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(9) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255−261.
(10) Gavin, A.-C.; Bosche, M.; Krause, R. et al. Nature **2002**, *415*, 141−147.
(11) Ho, Y.; Gruhler, A.; Heilbut, A. et al. Nature **2002**, *415*, 180−183.
(12) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(13) Edwards, A. M.; Arrowsmith, C. H.; Christendat, D.; Dharamsi, A.; Friesen, J. D.; Greenblatt, J. F.; Vedadi, M. *Nat. Struct. Biol.* **2000**, *7* Suppl., 970−972.
(14) Figeys, D. *Anal. Chem*. **2002**, 74, 413A−419A.
(15) Anderson, L.; Seilhamer, J. *Electrophoresis* **1997**, *18*, 533−537.
(16) Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1998**, *19*, 1853−1861.
(17) Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. **1999**, *19*, 1720−1730.
(18) Miklos, G. L. G.; Maleszka, R. *Proteomics* **2001**, *1*, 169−178.
(19) Anderson, L.; Anderson, J. *Proteomics* **2000**, *1*, 1−13.
(20) Bratt, C.; Lindberg, C.; Lindberg, C.; Marko-Varga, G. J. *Chromatogr. A* **2001**, 909, 279−288.
(21) Fey, S. J.; Larsen, P. M. *Curr. Opin. Chem. Biol.* **2001**, 5, 26−33.
(22) Malmstrom, J.; Bratt, C.; Lindberg, H.; Lindberg, C.; Burns, J.; Fey, S.; Mose Larsen, P.; Wieslander, E.; Särnstrand, B.; Marko-Varga, G. *Mol. Cell. Proteomics*, submitted, 2003.
(23) Veenstra, T. D. *Adv. Protein Chem*. **2003**, *65*, 161−194.
(24) Malmstrom, L.; Malmstrom, J.; Marko-Varga, G.; Westergren-Thorsson, G. *J. Proteome Res.* **2002**, *1*, 135−138.
(25) Unger, K. K.; Racaityte, K.; Wagner, K., T.; Edholm, L.-E.; Marko-Varga, G.; Bischoff, R.; Miliotis, J. *High Res. Chromatogr.* **2000**, *23*, 259−265.
(26) Machtejevas, E.; John, H.; Wagner, K.; Ständker, L.; Marko-Varga, G.; Forssmann, W. G.; Bischoff, R.; Unger, K. K., *J. Chromatogr. B* **2003**, submitted.
(27) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem*, **2001**, *70*, 437−473.
(28) Link, A. J.; Eng, J.; Schietz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garwik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(29) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269−295.
(30) Smith, R. D.; Pasa-Tolic, L.; Lipton, M. S.; Jensen, P. K. *Electrophoresis* **2001**, *9*, 1652−1668.
(31) Wu, S.-L.; Choudhary, G.; Ramström, M.; Bergquist, J.; Hancock, W. S. *J. Proteome Res.* **2003**, *2*, 383−393.
(32) Malmström, J.; Larsen, K.; Malmström, L.; Tufvesson, E.; Parker, K.; Marchese, J.; Williamsson, B.; Patterson, D.; Martin, S.; Westergren-Thorson, G.; Juhasz, P.; Marko-Varga, G. *J. Proteome Res.*, in press.
(33) Bodnar, W. M.; Blackburn, R. K.; Krise, J. M.; Moseley, M. A. *J. Am. Soc. Mass Spectrom*. **2003**, *14(9)*, 971−979.
(34) McDonald, W. H.; Yates, J. R, 3rd. *Curr. Opin. Mol. Ther.* **2003**, *5*(3), 302−309.
(35) McDonald, W. H.; Yates, J. R. 3rd *Dis. Markers* **2002**, *18(2)*, 99−105.
(36) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. *J. Proteome Res.* **2002**, *1*, 47−54.
(37) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotech*. **2002**, *20*, 512−515.
(38) Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269−295.
(39) Giaever, G.; Chu, A. M.; Ni, L. et. al. *Nature* **2002**, *418*(6896), 387−391.
(40) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. Nature. 2003 Oct 16; 425(6959): 737−41.
(41) Huh, W. K.; Falvo, J. V.; Gerke, L. C.; Carroll, A. S.; Howson, R. W.; Weissman, J. S.; O'Shea, E. K. *Nature* **2003**, *425*(6959), 686−691.
(42) Flory, M. R.; Griffin, T. J.; Martin, D.; Aebersold, R. *Trends Biotechnol.* **2002**, 12 Suppl., S23−29.
(43) Hanash. S. H.; Bobek, M. P.; Rickman, D. S.; Williams, T.; Rouillard, J. M.; Kuick, R. Puravs. *Proteomics* **2002**, 2, 69−75.
(44) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*(10), 994−999.
(45) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*(5), 512−515.
(46) Goshe, M. B.; Smith, R. D *Curr. Opin. Biotechnol.* **2003**, *14*(1), 101−109.
(47) Kalghatgi, K.; Horvath, C. S. *J. Chromatogr*. **1987**, *398*, 335.
(48) Molnar, I.; Horvath, Cs. *J. Chromatogr.* **1977**, *143*, 39−48.
(49) Ma, S.; Kalman, F.; Kalman, A.; Thunecke, F.; Horvath, C. *J. Chromatogr A.* **1995**, *711*, 67−82.
(50) Laurell, T.; Marko-Varga, G. *Proteomics* **2002**, *2*, 345−351.
(51) Marko-Varga, G.; Nilsson, J.; Laurell, T. *Electrophoresis* **2003**, *24*, 3521−3532.
(52) Walcher, W.; Franze, T.; Weller, M. G.; Pöschl, U.; Huber, C. G.; *J. Proteome. Res.* **2003**, in pres.
(53) Vollmer, M.; Nägele, E.; Hörth, P. *J. Biomol. Techn*ol. **2003**, *14*, 128−132.
(54) Rocklin, R. D.; Ramsey, R. S.; Ramsey, J. M. *Anal. Chem*. **2000**, *72*, 5244−5249.
(55) Michels, D. A.; Hu, S.; Schoenherr, R. M.; Eggertson, M. J.; Dovichi, N. J. *Mol. Cell. Proteomics* **2002**, *1*, 69−74.
(56) Tabuchi, T.; Yasuhiro, K.; Nakamura, K.; Baba, Y.; *J. Proteomics Res.* **2003**, *2*, 431−435.
(57) Templin, M. F.; Stoll, D.; Schwenk, J. M.; Pötz, O.; Kramer, S.; Joos, T. O. *Proteomics* **2003**, *3*, 2155−2166.
(58) Seong, S.-Y.; Choi, C.-Y. *Proteomics* **2003**, *3*, 2176−2189.
(59) Wulfkuhle, J. D.; Aquino, J. A.; Calvert, V. S.; Fishman, D. A.; Coukos, G.; Liotta, L. A.; Petricoin III, E. F. *Proteomics* **2003**, *3*, 2085−2090.
(60) Zhu, H.; Snyder, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 40−45.
(61) Robinson, W. H.; DiGennaro, C.; Hueber, W. et al. *Nat. Med.* **2002**, *8*, 295−301.
(62) Rees, J. *Science* **2002**, *296*, 698−700.
(63) Thomas, L. *New Engl. J. Med*. **1971**, *284*(19), 1082−1083.
(64) Reboul, J.; Vaglio, P.; Rual, J. F. et. al. *Nat. Genet*. **2003**, *34*(1), 35−41.
(65) Mawuenyega, K. G.; Kaji, H.; Yamuchi, Y.; Shinkawa, T.; Saito, H.; Taoka, M.; Takahashi, N.; Isobe, T. *J. Proteome Res.* **2003**, *2*(1), 23−35.

(66) Uetz, P. et al. *Nature* **2000**, *403*, 623−627.
(67) Florens, L.; Washburn, M. P.; Raine, J. D. et al. *Nature* **2002**, *419*(6906), 520−526.
(68) Giot, L.; Bader, J. S.; Brouwer, C. et al. *Science* **2003**, *302(5651)*, 1727−1736.
(69) Li, S.; Armstrong, C. M.; Bertin, N. et al. *Science* **2004**, *303(5657)*, 540−543.
(70) Camon, E.; Magrane, M.; Barrell, D.; Binns, D.; Fleischmann, W.; Kersey, P.; Mulder, N.; Oinn, T.; Maslen, J.; Cox, A.; Apweiler, R. *Genome Res.* **2003**, *13*(4), 662−672.
(71) Pawson, T.; Nas, P. *Science* **2003**, *300*, 445−452.
(72) Pandey, A.; Podtelejnikov, A. V.; Blagoev, B. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97(1)*, 179−184.
(73) Pandey, A.; Andersen, J. S.; Mann, M. *Sci. STKE.* **2000**, *37*, PL1.
(74) Zhoud, Z.; Licklider, L. J.; Gygi, S. P.; Reed, R. *Nature* **2002**, *419(6903)*, 182−185.
(75) Jorissen, R. N.; Walker, F.; Pouliot, N.; Garrett, T. P.; Ward, C. W.; Burgess, A. W. *Exp. Cell. Res.* **2003**, *284*(1), 31−53.
(76) Blagoev, B.; Kratchmarova, I.; Ong, S. E.; Nielsen, M.; Foster, L. J.; Mann, M. *Nat. Biotechnol.* **2003**, *21*(3), 315−318.
(77) Espina, V.; Dettloff, K. A.; Cowherd, S.; Petricoin, E. F. 3rd; Liotta, L. A. *Expert Opin. Biol. Ther.* **2004**, *1*, 83−93.
(78) Bonner, R. F.; Emmert-Buck, M.; Cole, K.; Pohida, T.; Chuaqui, R.; Goldstein, S.; Liotta, L. A. *Science* **1997**, *21*, 1482−1483.
(79) Michener, C. M.; Ardekani, A. M.; Petricoin, E. F., 3rd; Liotta, L. A.; Kohn, E. C. *Cancer Detect Prev.* **2002**, *26*(4), 249−255.
(80) Fend, F.; Raffeld, M. *J. Clin. Pathol.* **2000**, *53*(9), 666−672.
(81) Petricoin, E.; Hackett, J.; Lesko, L.; Puri, R. K.; Gutman, S.; Chumakov, K.; Woodcock, J.; Feigal, D. W.; Zoon, K. C.; Sistare, F. D. *Nat. Gen.* **2002**, *32*, 474−479.
(82) Issaq, H. J.; Conrads, T. P.; Prieto, D. A.; Tirumalal, R.; Veenstra, T. D. *Anal. Chem.* **2003**, 149A−155A.
(83) Frank, R.; Hargreaves, R. *Nat. Rev. Drug Discov.* **2003**, *7*, 566−580.
(84) Savage, D. R. *Nat. Rev. Drug. Discuss.* **2003**, *2*, 937−939.
(85) Lesko, L. J.; Salerno, R. A.; Spear, B. B. *J. Clin. Pharmacol.* **2003**, *43(4)*, 342−358.
(86) Service, R. F. *Science* **2003**, *302(5649)*, 1316−1318.
(87) Edelman, G. M.; Gally, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*(24), 13 763−13 768.
(88) Plymoth, A.; Löfdahl, C.-G.; Ekberg-Jansson, A.; Dahlbäck, M.; Fehniger, T. E.; Marko-Varga, G. *Abstracts of 99th ATS Meeting*; Seattle, Washington, USA, 2003, P. A741.
(89) Zhang, H.; Li X−J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(90) Vazquez, A.; Flammini, A.; Maritan, A.; Vespignani, A. *Nat. Biotechnol.* **2003**, *21*, 697−702.
(91) Patterson, S. D. *Nat. Biotechnol.* **2003**, *21*, 221−222.
(92) Jensen, L. J.; Gupta, R.; Staerfeldt H−H.; Brunak, S. *Bioinformatics* **2003**, *19*, 635−642.
(93) Jerne, N. K. *Australas Ann. Med.* **1969**, *18*(4), 345−348.

PR049958+