Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders

Proteomics approaches to the systems biology of cardiovascular diseases

Thomas A. Drake[1],* and Peipei Ping[†,§]

Departments of Pathology and Laboratory Medicine,* Physiology,[†] and Medicine,[§] University of California Los Angeles, Los Angeles, CA

Abstract   Proteins play a central role in a systems view of biologic processes. This review provides an overview of proteomics from a systems perspective.▯▯ We survey the key tools and methodologies used, present examples of how these are currently being used in the systems biology context, and discuss future directions.—Drake, T. A., and P. Ping. **Proteomics approaches to the systems biology of cardiovascular diseases.** *J. Lipid Res.* 2007. **48:** 1–8.

As defined in the introduction to this series, a "systems level" approach implies the examination of an ensemble of multiple elements, as opposed to individual elements on their own. Our review contributes a proteomics perspective. We will survey the key methodologies used, present examples of how these are currently being used in the systems biology context, and discuss future directions.

As a category of "omics" molecules, proteins play a central role in a systems view of biologic processes. Proteins make up a major portion of cell structural molecules and are the effector molecules for all cell functions, either directly or indirectly through their actions on metabolites and other nonprotein substrates. The complexity of the proteome far exceeds that of the genome and transcriptome as a result of a range of posttranslational modifications and interactions with other molecules. Subcellular compartmentalization and a potential multiplicity of interaction partners expand this further.

The complexity of the proteome and our current inability to globally interrogate it to the extent we can the transcriptome significantly limits the scope of current proteomics studies. In contrast to the transcriptome, in which the chemistry of nucleic acids allows a common technological approach that is capable of approaching global assessment, the technologies used for proteomic investigations are numerous and varied, often focused on a specific protein subset. Any one "high-throughput" method typically detects on the order of a few thousand proteins at most, and generally far fewer. Because these technological issues frame the questions that can be asked and the data collected, we will provide a summary of methodologies at the start (**Fig. 1**).

TOOLS USED IN PROTEOMICS

Mass spectrometry is nearly synonymous with proteomics because it is an effective tool and is widely available for protein identification in an undirected (i.e., without prior knowledge) manner. The classical principle of mass spectrometry is followed in most of the current instruments: molecules undergo ionization and are accelerated toward a detector in a vacuum environment using magnetic forces. The time to detection ("time of flight") is proportional to both the mass and the charge of the molecule, such that the time of flight is approximately defined by the mass-to-charge ratio, typically designated as m/z, which is the primary measurement obtained in an analysis. The data from an analysis obtained from one sample that has undergone ionization can be displayed as a "spectrum," which is a plot of the m/z ratio on the x axis versus the level of intensity as measured by the detector on the y axis. Intensity roughly reflects the relative number of ionized molecules detected at a given time. In this graphical format, a peak reflects the presence of one or more molecules. The resolving power of a mass spectrometer varies inversely with the m/z ratio, and there may be significant differences in overall resolving power among different instruments. Regardless of instrumentation, a

[1] To whom correspondence should be addressed.
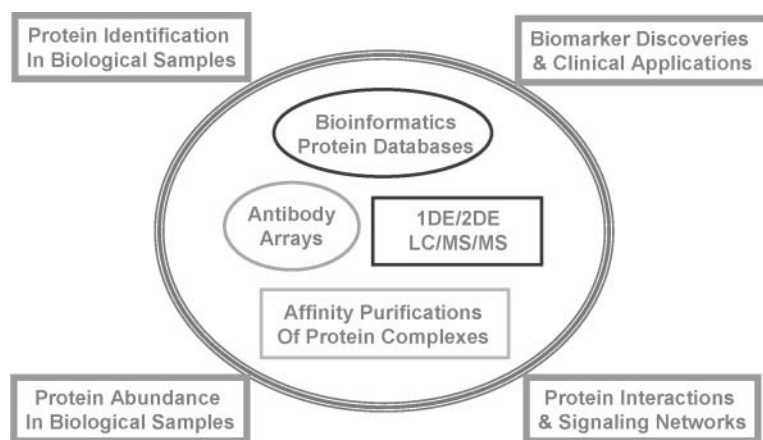   e-mail: tdrake@mednet.ucla.edu

Fig. 1. Diagram of the current core proteomic technologies as well as the four key applications they drive in advancing biology and medicine. 1DE/2DE LC/MS/MS, one- and two-dimensional liquid chromatography combined with tandem mass spectroscopy.

given m/z is not unique to a given molecule, so additional steps and other information are necessary to identify a protein. Because resolution is usually poor above m/z ~10,000, proteins are typically hydrolyzed to peptides before analysis. Liquid chromatography combined with tandem mass spectroscopy (LC-MS/MS) has the capability to capture ions at a particular m/z and perform sequential MS analysis of fragmented products of these ions, which allows for peptide identification. Other protein separation methodologies typically use mass spectroscopy as a final step for identification. There are many variations on instrumentation and sample preparation and their use in various applications. A number of recent reviews discuss these, and we will describe some in the context of their use below (1–5).

Two-dimensional gel electrophoresis has been used for many years for the separation of intact proteins. Samples typically undergo sequential electrophoresis, first using isoelectric focusing to separate by charge, followed by polyacrylamide gel electrophoresis to separate by size. Separated proteins are visualized by imaging after staining with fluorescent dye. Prelabeling of samples with dye allows the use of several different dyes for different samples and simultaneous analyses of several samples on one gel (6). This allows for internal controls and more accurate quantitation. The proteins composing the resulting ''spots'' can be identified by physically removing a plug of the gel containing the protein, hydrolyzing it with trypsin, and then performing tandem MS as described above. For a given protein, multiple spots are often visualized, reflecting charge differences attributable to posttranslational modifications. Resolution and sensitivity are improved for complex samples by prior fractionation or by removal of high-abundance proteins, such as albumin and immunoglobulins, when analyzing plasma.

Protein microarray refers to a general platform for capturing proteins by immobilizing potential binding partners on a solid surface, followed by detection and identification of the bound protein (7). Binding partners are typically predefined and could be DNA that includes potential transcription factor binding sequences, an array of antibodies for a designated set of targets, or an array of proteins, etc. Miniaturization of platforms for efficient high-throughput use is an active area of development. Similarly, microfluidic in-solution assay systems are being developed that allow for multianalyte high-throughput use.

Yeast two-hybrid systems assess possible paired protein-protein interactions on a large scale, sometimes referred to as the ''interactome'' (8). These use specialized yeast libraries carrying cloned protein coding open reading frames (ORFs) that are expressed as fusion proteins with transcription factor elements whose activity is necessary and in principle occurs only when there is physical interaction between the two expressed ORFs. The identity of the ORF is determined by sequencing of PCR amplification products of the ORFs of positive sets. These are prone to significant rates of false-positive and false-negative results in any one experiment. Methodological variations on the fusion protein elements and the mechanism of selection of positively interacting clones have been developed.

Antibody development is an important enabling tool set being developed that can be applied in a number of different settings. The ultimate goal is to generate antibody reagents available for all mammalian proteins. Their applications in conjunction with protein microarrays, imaging technologies, and other biochemical tools will enable the illustration of the mammalian protein atlas in individual tissues/organs. In addition, they may be used in conjunction with tissue microarrays to define patterns of cellular expression for signaling networks and pathways, with two-dimensional gel electrophoresis for protein identifications, and as a prefractionation method (e.g., immunoaffinity-based purifications) for subsequent MS analyses.

Computational tools for proteomic sciences contain at least two major elements: one is bioinformatic analyses of cellular proteins, and the second is the construction of protein databases available to the public domain (9–12). The development of both of these tools in proteomics has benefited from achievements made in genomic research. For example, in most uncharacterized proteins, putative amino acid sequences may be predicted on the basis of their DNA data. Leader peptide sequences, if present, may suggest what compartment the protein likely resides in, such as mitochondria, or secretion, and various domain motifs can be indicative of transmembrane proteins or DNA binding functions, for example (13). Comparative

analyses across species for conserved proteins may also be helpful. There is an increased consensus in the proteomic community for protein data standardization; this effort has gained support and input from investigators, proteomic instrumentation vendors, journals, and funding agencies. Protein databases designated for protein data archiving and analyses are made available to the public and can be readily used for comparison analyses among different species, organs, organelles, cellular networks, and pathways.

QUESTIONS PROTEOMICS CAN ANSWER

Because of the complexity of the proteome "universe," there is a broad scope to research in this field. Rather than attempt to be comprehensive, we will address here five basic questions one might have from a systems perspective concerning a given tissue or cell: What are the proteins that are present? How much of each is present under different conditions? Which of these interact with one another? What are the functional states of proteins whose activity may not be reflected by the amount present? How can this information be used clinically?

Addressing the exploratory question "What proteins are present in a biologic compartment?" has been and continues to be a major focus of proteomic study. The starting point of any systems-based approach is the definition of the "parts list," and from a global perspective, this is far from complete for any tissue or disease. Tissue or cell samples may be analyzed in toto or undergo varying levels of cellular or subcellular fractionation before proteomic analysis (14). In the latter situation, investigators are often interested in a specific compartment or organelle, such as membrane proteins or mitochondria. When specific fractions are studied, it is important to recognize that "pure" fractions are impossible to obtain and that additional steps are necessary to validate localization, such as immunolocalization. Because this can usually be done on only a subset of proteins, there are always some false-positives in any given study.

Data from comprehensive analyses of all subcellular fractions are limited, but one such study published recently is informative as an example of protein profiling of multiple fractions and organelles (15). It demonstrates the extent of analysis involved for just one tissue and condition and is also representative of a common MS methodology used, LC-MS/MS. In preparatory steps, mouse liver homogenates underwent separation and gradient centrifugation, yielding a total of 32 fractions. Each fraction was digested to completion with trypsin before liquid chromatographic fractionation. Each LC fraction underwent MS using a linear ion trap/Fourier transform hybrid mass spectrometer (an instrument with very high resolution), and the five most abundant peaks from each spectra underwent fragmentation and MS for identification. More than 22,000 peptides were identified, representing just <2,200 unique proteins. Approximately 1,400 of these were able to be localized to one or more of 10 compartments using an analytical approach termed protein correlation profiles, which uses marker proteins to help assign compartment membership. These ranged from 50 in proteasomes to >300 in recycling endosomes. The extent of effort involved (mostly analysis time) was such that most data were obtained from one liver, emphasizing that detailed proteomic analyses remain extremely time- and effort-intensive relative to gene expression analyses and sample a smaller fraction of the total proteins expected to be present. Nevertheless, data from a number of studies using various methodologies have accumulated concerning the protein complement of various tissues and organelles relevant to cardiovascular disease, including studies of cultured endothelial and smooth muscle cells, as reviewed recently (16, 17). The plasma lipoproteins can also be considered a compartment of sorts, and several recent studies have analyzed these using the methods described above (18–21).

A caveat for this and subsequent discussion is that tissue samples are always complex with regard to the cell types present. Technical considerations preclude analysis of single cells, except by immunostaining methods. One cell type may compose the bulk of tissue, as in myocardium or liver, but in pathologic settings significant numbers of inflammatory cells may be present or there may be a change in the normal proportion of parenchymal to stromal cells (e.g., cardiomyocytes to fibroblasts). The analysis of atherosclerotic lesion material is perhaps an extreme example of this (22). The use of isolated or cultured cells addresses the complex cell type issue but raises other concerns, such as relevance to the in vivo setting. An intriguing approach that may find application in complex tissues such as the vessel wall is the use of MS to obtain spectra directly from intact tissue sections, termed "tissue profiling" (23).

A second important question is "What proteins change in concentration in response to physiological or pathological perturbations?" The ability to make sensitive quantitative measurements is fundamental to the essential nature of systems biology investigation, which is to assess system elements over a range of perturbations. Unfortunately, most standard undirected proteomic methods are inherently semiquantitative at best. Although abundance differences look obvious in a given gel or MS spectrum, obtaining accurate measures over multiple samples analyzed at different times is difficult because of numerous run-to-run sample preparation and technical variations. A sense of protein abundance can be gained simply from the frequency of detecting a given protein in a series of samples. High-abundance proteins will be identified many more times than low-abundance proteins. However, significant effort has been made to develop truly quantitative approaches, and these are available. Most of these assess relative abundance between test and control samples using differential labeling, but one recent method allows absolute quantitation.

When using two-dimensional gel methods, normalization for technical differences influencing spot location and intensity across multiple gels is the major problem for accurate quantitation. Reasonable quantitation has been

obtained by careful attention to conditions and using normalization methods based on total spot intensities. However, better relative quantitation can be achieved using the differential in-gel expression (DIGE) technique (6). DIGE addresses this by creating a control composed of a pool of aliquots from all of the experimental samples that is run with every gel. Different fluorescent dyes are used to label the proteins in test and control samples, typically a Cy3 or Cy5 label for each individual test and control sample and Cy2 for the pooled sample. In a given gel, for each spot the results of the test samples are expressed as a ratio to the pool. Coefficients of variation of 10% or less can be obtained with this method. A related method described recently is the intact protein analysis system (IPAS), in which paired samples are labeled with Cy3 or Cy5 and then subjected to extensive fractionation based on charge (by isoelectric focusing), hydrophobicity (reverse-phase HPLC), and mass (SDS-PAGE) (24, 25).

Among MS approaches, direct quantitation of peak intensities can be used when carefully performed to minimize sample handling differences and including standards and controls for instrument calibration and normalization. However, more accurate quantitation is achievable using differential sample labeling methods. For MS analyses, differential labeling yields peaks with predictable differences in m/z ratio, allowing one to obtain a ratio of experimental to control intensity for any given peak. Such labels that change peptide mass can be applied exogenously, as in the Isotope Coded Affinity Technology method, or endogenously by providing a heavy isotope in culture medium or diet, which becomes incorporated into proteins (26–28). Reagents for labeling are relatively expensive, and additional analytical steps are needed to match paired peaks for data analysis. Recently, a method was developed that allows absolute quantitation using synthesized labeled peptides (29, 30). These are added to test samples in known amounts at an early stage of sample separation and therefore are subjected to the same processing and technical variations. Because an absolutely defined amount has been added to a sample, the ratio of test to control peptide intensity can provide an absolute quantitation. Again, the reagents are expensive, and one needs to determine beforehand which proteins one wishes to measure and prepare reagents for each.

Answering "Which proteins interact physically with one another?" is a central aim of much proteomics research. Although important relationships may exist between proteins in the absence of direct physical interactions, the converse is not likely to be true. Most important cellular processes occur with at least the transient association of multiple proteins, and in many instances associations are quite stable and required for function, such as the majority of the electron transport chain in mitochondria or the structural molecules composing organelles.

There are two general approaches for determining protein-protein interactions. A "local" approach is centered on an individual protein of interest as a "capture" protein and isolates the complex of proteins that are physically attached to that. The complex of proteins is then analyzed by MS, and the individual proteins present are identified. One can then extend this local set of interacting proteins by repeating the process using one of the newly identified proteins as the capture protein. Through repetition of this sequence, a network of interactions can be constructed. The strength of this approach is that in most configurations it examines protein interactions as they occur in the cell. However, it should be noted that although such constructs are studied in living cells, their level of expression may be outside the physiologic range and thus subject to various biases. A variety of methods and configurations have been developed for this approach (31–33). For example, some use expression of transfected constructs containing the primary protein as a fusion with elements that facilitate its capture, whereas others evaluate endogenous proteins by treating cells or tissue to cause cross-linking, followed by pull-down with antibodies. We present examples of this approach from our own work below.

A "global" approach is represented by yeast two-hybrid analyses, in which all possible pairs of interacting proteins are examined. By the nature of the system, the protein interactions detected are not taking place in their natural setting within a cell, so a positive finding indicates the potential for interaction but not proof that the interaction occurs in a natural environment. Dynamic changes that might occur in varying physiological or disease states cannot be assessed. These approaches are also subject to relatively high false-negative and false-positive results in any one experiment, but the finding of consistent results over multiple experiments reduces these. The usefulness of this type of approach is that it examines a much larger set of proteins than is possible otherwise and suggests potential interactions among sets of proteins that can be validated by other means. A recent article reports such an analysis that is the largest to date of human proteins. It used a set that included ~7,200 distinct human protein-encoding genes and identified ~2,800 high-confidence interacting pairs, almost 80% of which in sampling could be validated by alternative techniques (34).

Assessing "What is the functional state of a protein?" is a major challenge. In the absence of a direct functional assay, detection of several types of common structural modifications to proteins can be used to infer functional differences. The best known is phosphorylation, which may confer enzymatic or binding activity to an otherwise unreactive protein. Modifications made by the specific addition of ubiquitin or related small molecules such as SUMO (small ubiquitin-related modifier) also confer altered protein properties, and are highly regulated in cells. Ubiquitination targets proteins for degradation while SUMOylation affects subcellular localization and stabilization of proteins. Glycosylation is a common posttranslational modification but does not have such specific functional implications. These posttranslational modifications occur at selective sites on proteins and confer an altered mass; they also establish a new epitope. The latter has been used to develop antibody reagents that can distinguish phosphorylated from nonphosphorylated states

for a number of important proteins, especially those involved in signaling pathways. These have been configured into array formats for simultaneously examining many at one time. MS can be used in an undirected manner to identify phosphorylated proteins (phosphoproteins). For a recent example, see Wolf-Yadlin et al. (35). This is done in conjunction with preanalysis steps to enrich for these proteins. The most well established is immobilized metal affinity chromatography. Trypsinized homogenates of cells or tissue are applied, and the metal chelating surface selectively binds phosphorylated peptides from the digest, which then can be identified by MS. Ubiquitinated peptides from a cell digest can be enriched for using immunoaffinity chromatography with antibodies to ubiquitin, or in systems in which a recombinantly modified ubiquitin is expressed wherein a capture tag has been added. MS is then used to identify the peptides and verify the presence of the ubiquitin modification.

Another approach toward assessing the functional states of proteins relates to active versus inactive enzymes, in which the inactive form may exist as a zymogen or as an intact protein bound by an inhibitor. These may not be distinguished by the standard methods described above. An approach referred to as "activity-based protein profiling" uses active site-directed probes to capture the active enzymes, but not the inactive or inhibited forms, for identification by MS or other means (36–38).

"How can proteomic studies advance cardiovascular biology and inform clinical medicine?" is a focus of those interested in using this set of tools to advance the fundamental knowledge of biology and/or to identify potential "biomarkers" that may be indicative of a diseased phenotype.

Advancing our understanding of biology and disease by any of the above means benefits medicine. However, finding proteins that reflect health/disease state or risk and are present in noninvasively accessible samples from people is a special challenge. Biomarkers may be sought in a directed manner based on the kinds of studies described above in cells or tissue, followed by developing the means to assess their presence in blood, urine, saliva, or other body fluid, typically by immunoassay. The alternative approach is to directly analyze such samples by MS in an undirected search, and this is an active area of proteomics research.

Work in this area falls into two main categories, which are not mutually exclusive. The first encompasses studies to systematically define the proteomes of the different compartments, and the second encompasses studies searching for biomarkers for specific diseases. The plasma proteome has received the most attention, as it contains proteins originating from any part of the body that find their way into the circulation. The bulk of proteins are secreted, many from the liver, but nonsecreted proteins arising from cell "leakage" or death, matrix turnover, or pathologic processes are also present. They may be intact or proteolytic fragments of proteins and may circulate free or bound to carriers such as albumin or lipoprotein particles. These cover an extremely wide range of concentra-

tions, estimated to be as much as 10^{10}, which exceeds that of cells and tissues directly. The complexity and dynamic range makes defining the plasma proteome a challenge (39). The Human Proteome Organization (HUPO) has supported other research efforts in coordinating a systematic analysis of the plasma proteome by a number of laboratories using a range of techniques, as recently described and reported (40–42). A composite list of reliable identifications has been compiled and is publicly available. This identifies ∼3,000 proteins. The coordinated effort also led to very useful data concerning the technical aspects of the proteomic analysis of plasma (43). We discuss aspects of this project specifically relevant to cardiovascular disease below.

In the second category, much of the work has been directed at cancer biomarker discovery (44–46). All of the techniques described above have been used. To define a proteome, a thorough analysis that identifies as many proteins as possible is desired, and relatively few individual samples need to be analyzed. This approach can also be taken for biomarker discovery, making the assumption that a limited number of subjects from disease and reference populations are adequately representative, because such thorough analyses are extremely time- and effort-consuming. An alternative approach that has been applied trades thoroughness for high throughput (47–49). Data are analyzed for correlation with disease state, and identification of the proteins these represent is deferred until such an association is found. However, substantial effort is needed for the identification of peaks of interest, quantitation may be problematic, peak resolution is poor with low-resolution instruments, and the sensitivity is low for low-abundance proteins. Because proteins with molecular weight > 20,000 are not well detected, gel-based methods (DIGE or IPAS, described above) can provide complementary data, but these are not high throughput. Zhang and colleagues (50) have described a high-throughput quantitative approach using LC-MS/MS that is relatively sensitive. The success of these screening approaches has been debated, and newer ways of approaching biomarker discovery are clearly needed (44).

In either situation, it is important that identified biomarkers be understood at the pathophysiological level. Likewise, proteomic-based discoveries at this level can serve as starting points for biomarker discovery. Recent investigations using proteomic tools have offered novel information of cellular organelles in cardiovascular tissues that would not have been made available in such a timely manner with conventional biochemical approaches (16, 17). These biological advancements are made in vascular endothelial cells and in cardiac cells. Detailed delineation of organelle proteins coupled with convincing target validation leads to a full characterization of vascular endothelium-based caveolae, cardiac proteasome complexes, cardiac mitochondria, and cardiac contractile apparatus (reviewed in 16). The illustration of a protein atlas in these organelles will provide the much needed fundamental data essential for future investigations aimed at exploring their functional roles in cardiovascular diseases.

A last point about clinical biomarker studies is that consistency of sample collection techniques, processing, and storage is critical for reliable results to be obtained. This can be challenging in the clinical setting, and the samples collected in the course of the usual clinical study typically are inadequate for proteomic analyses when the latter were not considered at the start.

PROTEOMICS AND SYSTEMS BIOLOGY

The field of proteomics has made tremendous contributions to biomedical science overall, and it has specifically contributed to the developing systems perspective of biology and disease. The following three areas have been particularly informed by proteomics based research: characterization of biological systems as scale-free networks; elucidation of local modules or pathways; and disease biomarker development.

Protein interaction networks are scale-free

Systems biology attempts to characterize the structure and dynamics of interactions among elements of cells, tissue, and organisms. As presented in the introduction to this series and elaborated on by earlier reviews, biological systems are networks of interacting elements, and overall, these have a scale-free structure. Analyses based on "global" proteomic pair-wise interaction data obtained from various organisms have contributed significantly to the apparent universality of this observation (34, 51, 52). These networks are constructed with the proteins being the "nodes" and an observation of interaction being an "edge." Although not of apparent immediate relevance to cardiovascular disease per se, this general observation has important implications. The scale-free nature of the protein interaction network indicates that a limited number of proteins have a large number of interactions and function as "hubs." In general, hubs in biological networks are critical to overall network organization and function. As with gene expression-based networks, hub proteins are more frequently essential to organism survival than are nonhub proteins. A second characteristic of such networks is that they are composed of sets of "modules," groups of proteins that have significant numbers of within-group interactions compared with interactions outside the group. As physical interaction implies a functional consequence, such modules suggest that a particular group of proteins interact for a common function (or set of functions). Thus, protein interaction networks provide an indication of which proteins are more likely to be critical for overall functioning and indicate groups of proteins sharing common functions. Such information can guide more specific studies, as discussed below.

Proteomic studies are critical to elucidating functional modules

A systems-oriented paraphrasing of the saying "all politics is local" would be "all biology is local." Regardless of the high-level overall functioning of a cell, tissue, or organism, the nuts-and-bolts work happens at a local level among a set of closely interacting molecules. It is the summation and interactions of the effects of these smaller units that determine the direction of the entity as a whole, with the potential for developing the emergent properties of the larger system, as described by Weiss et al. (52a) in an earlier review in this series. One can envision the overall network as the highest level, or most abstract, view, with progressively smaller subsets of interacting molecules as modules, each with subnetworks that are dedicated to a specific functional process. From this perspective, the elucidation of these dedicated interacting subnetworks is key to defining the basic functional biochemical units of cells and tissues.

Our work in elucidating the signaling modules essential for protection of the heart against myocardial ischemic injury is illustrative in this regard (53, 54). In this line of investigation, a specific signaling kinase, protein kinase Cε (PKCε), was first characterized to be a key signaling node in cardiac cell cytoprotection. Evidence from a number of laboratories document that the activation of PKCε is a shared common signaling event in cytoprotection of the heart in species such as mouse, rat, rabbit, and human. In subsequent analyses using a functional proteomic approach, we identified a pool of candidate proteins that may serve as key associating partners of PKCε; these proteins form functional multiprotein complexes with PKCε. Using a combined genetic and proteomic approach, we identified several signaling modules (e.g., PKCε-Lck, PKCε-Akt, and PKCε-Bmx) as key signaling complexes leading to the manifestation of a cardioprotective phenotype.

PROTEOMICS-BASED STRATEGIES FOR DISEASE BIOMARKER DETECTION ARE BASED ON A SYSTEMS BIOLOGY PERSPECTIVE

Viewing disease processes from a systems perspective has changed the perspective on diagnostic testing. Current tests have been developed and largely used in a stand-alone manner. This necessitates a high level of sensitivity and specificity that may be difficult to achieve for both biologic and technical reasons. Strategies for test development now look toward identifying an optimal panel of analytes. Any one of them may have modest sensitivity and specificity, and among patients the pattern of activities may be different, yet together good predictive value is achieved (46). Computational algorithms can influence the apparent performance of such tests, and how these will be evaluated by the Food and Drug Administration for clinical use remains to be seen.

A second way that perceptions have changed is in the appreciation of how very limited our current set of analytes is relative to the number possible (55). Work toward defining the plasma proteome has increased the number of known proteins by more than an order of magnitude, and it still has a long way to go. We have analyzed the data obtained from the cooperative HUPO Plasma Proteome Project for proteins of likely relevance to cardiovascular

disease (56). Both cardiac tissue-born proteins (released into the plasma) and plasma-born proteins are included in this annotation. These proteins may be categorized as in distinct functional groups, including molecular markers of inflammation and/or cardiovascular disease, vascular and coagulation, signaling, growth and differentiation, cytoskeletal, transcription factors, channels/receptors, heart failure, and remodeling. Importantly, our analyses of the peptide per protein ratio for LC-MS/MS identifications display group-specific trends, corroborating the functional classification of these plasma proteins. The incorporation of proteomic analyses into clinical trials, as described for the CardioGene Study (57), will greatly expand our knowledge in this area in the long run.

CONCLUSIONS AND FUTURE DIRECTIONS

Integration of proteomics data with other "omics" data sets to model cell functions at different levels will bring the real potential of systems biology to bear on elucidating disease processes (58–61). Formal modeling is still in the earliest stages, but investigators are increasingly combining both proteomic and expression analyses to obtain deeper understanding of cell behavior than either analysis alone could provide (62–69). Increasingly, other data sets from metabolomic and transcriptional regulatory studies (among others) will be compared and integrated, as Gerstein and colleagues have shown in yeast (52). A further dimension is the incorporation of genetic variation (70).

Technological advances will undoubtedly be important drivers in the future, as they have been. Most proteomics studies are labor- and time-intensive, requiring a high level of expertise, and are relatively expensive, all of which limit the widespread use of these approaches. With time, the technologies can be expected to become more accessible, as happened with gene expression microarray technologies. However, as these become more and more used, a key issue of proteomic investigation that requires attention is that these unbiased analyses often only provide a way to a narrowed search: they may not lead to a definitive identification of the true targets. It cannot be overemphasized that subsequent studies using biochemical and functional tools to verify the proteins characterized through proteomic studies are of critical importance and are the integral part of successful proteomic investigations.■

REFERENCES

1. Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. *Nature.* **422:** 198–207.
2. de Hoog, C. L., and M. Mann. 2004. Proteomics. *Annu. Rev. Genomics Hum. Genet.* **5:** 267–293.
3. Domon, B., and R. Aebersold. 2006. Mass spectrometry and protein analysis. *Science.* **312:** 212–217.
4. Ferguson, P. L., and R. D. Smith. 2003. Proteome analysis by mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **32:** 399–424.
5. Glish, G. L., and R. W. Vachet. 2003. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov.* **2:** 140–150.
6. Van den Bergh, G., and L. Arckens. 2004. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. *Curr. Opin. Biotechnol.* **15:** 38–43.
7. Kingsmore, S. F. 2006. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat. Rev. Drug Discov.* **5:** 310–320.
8. Cusick, M. E., N. Klitgord, M. Vidal, and D. E. Hill. 2005. Interactome: gateway into systems biology. *Hum. Mol. Genet.* **14 (Spec. No. 2):** R171–R181.
9. Desiere, F., E. W. Deutsch, N. L. King, A. I. Nesvizhskii, P. Mallick, J. Eng, S. Chen, J. Eddes, S. N. Loevenich, and R. Aebersold. 2006. The PeptideAtlas project. *Nucleic Acids Res.* **34:** D655–D658.
10. Hsu, F., T. H. Pringle, R. M. Kuhn, D. Karolchik, M. Diekhans, D. Haussler, and W. J. Kent. 2005. The UCSC Proteome Browser. *Nucleic Acids Res.* **33:** D454–D458.
11. Martens, L., H. Hermjakob, P. Jones, M. Adamski, C. Taylor, D. States, K. Gevaert, J. Vandekerckhove, and R. Apweiler. 2005. PRIDE: the proteomics identifications database. *Proteomics.* **5:** 3537–3545.
12. Mishra, G. R., M. Suresh, K. Kumaran, N. Kannabiran, S. Suresh, P. Bala, K. Shivakumar, N. Anuradha, R. Reddy, T. M. Raghavan, et al. 2006. Human protein reference database—2006 update. *Nucleic Acids Res.* **34:** D411–D414.
13. Schneider, G., and U. Fechner. 2004. Advances in the prediction of protein targeting signals. *Proteomics.* **4:** 1571–1580.
14. Kislinger, T., B. Cox, A. Kannan, C. Chung, P. Hu, A. Ignatchenko, M. S. Scott, A. O. Gramolini, Q. Morris, M. T. Hallett, et al. 2006. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell.* **125:** 173–186.
15. Foster, L. J., C. L. de Hoog, Y. Zhang, Y. Zhang, X. Xie, V. K. Mootha, and M. Mann. 2006. A mammalian organelle map by protein correlation profiling. *Cell.* **125:** 187–199.
16. Mayr, M., J. Zhang, A. S. Greene, D. D. Gutterman, J. K. Perloff, and P. Ping. 2006. Proteomic based development of biomarkers in cardiovascular disease: mechanistic, clinical, and therapeutic insights. *Mol. Cell. Proteomics.* **5:** 1853–1864.
17. McGregor, E., and M. J. Dunn. 2006. Proteomics of the heart: unraveling disease. *Circ. Res.* **98:** 309–321.
18. Heller, M., D. Stalder, E. Schlappritzi, G. Hayn, U. Matter, and A. Haeberli. 2005. Mass spectrometry-based analytical tools for the molecular protein characterization of human plasma lipoproteins. *Proteomics.* **5:** 2619–2630.
19. Karlsson, H., P. Leanderson, C. Tagesson, and M. Lindahl. 2005. Lipoproteomics II: mapping of proteins in high-density lipoprotein using two-dimensional gel electrophoresis and mass spectrometry. *Proteomics.* **5:** 1431–1445.
20. Karlsson, H., P. Leanderson, C. Tagesson, and M. Lindahl. 2005. Lipoproteomics I: mapping of proteins in low-density lipoprotein using two-dimensional gel electrophoresis and mass spectrometry. *Proteomics.* **5:** 551–565.
21. Rezaee, F., B. Casetta, J. H. Levels, D. Speijer, and J. C. Meijers. 2006. Proteomic analysis of high-density lipoprotein. *Proteomics.* **6:** 721–730.
22. Mayr, M., Y. L. Chung, U. Mayr, X. Yin, L. Ly, H. Troy, S. Fredericks, Y. Hu, J. R. Griffiths, and Q. Xu. 2005. Proteomic and metabolomic analyses of atherosclerotic vessels from apolipoprotein E-deficient mice reveal alterations in inflammation, oxidative stress, and energy metabolism. *Arterioscler. Thromb. Vasc. Biol.* **25:** 2135–2142.
23. Caldwell, R. L., and R. M. Caprioli. 2005. Tissue profiling by mass spectrometry: a review of methodology and applications. *Mol. Cell. Proteomics.* **4:** 394–401.
24. Misek, D. E., R. Kuick, H. Wang, V. Galchev, B. Deng, R. Zhao, J. Tra, M. R. Pisano, R. Amunugama, D. Allen, et al. 2005. A wide range of protein isoforms in serum and plasma uncovered by a quantitative intact protein analysis system. *Proteomics.* **5:** 3343–3352.
25. Wang, H., S. G. Clouthier, V. Galchev, D. E. Misek, U. Duffner, C. K. Min, R. Zhao, J. Tra, G. S. Omenn, J. L. Ferrara, et al. 2005. Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. *Mol. Cell. Proteomics.* **4:** 618–625.
26. Goshe, M. B., and R. D. Smith. 2003. Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.* **14:** 101–109.
27. Ishihama, Y., T. Sato, T. Tabata, N. Miyamoto, K. Sagane, T. Nagasu, and Y. Oda. 2005. Quantitative mouse brain proteomics

using culture-derived isotope tags as internal standards. *Nat. Biotechnol.* **23:** 617–621.

28. Wu, C. C., M. J. MacCoss, K. E. Howell, D. E. Matthews, and J. R. Yates, 3rd. 2004. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* **76:** 4951–4959.

29. Ong, S. E., and M. Mann. 2005. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1:** 252–262.

30. Pan, S., H. Zhang, J. Rush, J. Eng, N. Zhang, D. Patterson, M. J. Comb, and R. Aebersold. 2005. High throughput proteome screening for biomarker detection. *Mol. Cell. Proteomics.* **4:** 182–190.

31. Forler, D., T. Kocher, M. Rode, M. Gentzel, E. Izaurralde, and M. Wilm. 2003. An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat. Biotechnol.* **21:** 89–92.

32. Gingras, A. C., R. Aebersold, and B. Raught. 2005. Advances in protein complex analysis using mass spectrometry. *J. Physiol.* **563:** 11–21.

33. Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods.* **24:** 218–229.

34. Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* **437:** 1173–1178.

35. Wolf-Yadlin, A., N. Kumar, Y. Zhang, S. Hautaniemi, M. Zaman, H. D. Kim, V. Grantcharova, D. A. Lauffenburger, and F. M. White. 2006. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* **2:** 54.

36. Barglow, K. T., and B. F. Cravatt. 2004. Discovering disease-associated enzymes by proteome reactivity profiling. *Chem. Biol.* **11:** 1523–1531.

37. Liu, Y., M. P. Patricelli, and B. F. Cravatt. 1999. Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA.* **96:** 14694–14699.

38. Okerberg, E. S., J. Wu, B. Zhang, B. Samii, K. Blackford, D. T. Winn, K. R. Shreder, J. J. Burbaum, and M. P. Patricelli. 2005. High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl. Acad. Sci. USA.* **102:** 4996–5001.

39. Jacobs, J. M., J. N. Adkins, W. J. Qian, T. Liu, Y. Shen, D. G. Camp, 2nd, and R. D. Smith. 2005. Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* **4:** 1073–1085.

40. Omenn, G. S. 2004. Advancement of biomarker discovery and validation through the HUPO Plasma Proteome Project. *Dis. Markers.* **20:** 131–134.

41. Omenn, G. S. 2004. International collaboration in clinical chemistry and laboratory medicine: the Human Proteome Organization (HUPO) Plasma Proteome Project. *Clin. Chem. Lab. Med.* **42:** 1–2.

42. Omenn, G. S., D. J. States, M. Adamski, T. W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B. B. Haab, R. J. Simpson, J. S. Eddes, et al. 2005. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics.* **5:** 3226–3245.

43. Omenn, G. S. 2004. The Human Proteome Organization Plasma Proteome Project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. *Proteomics.* **4:** 1235–1240.

44. Aebersold, R., L. Anderson, R. Caprioli, B. Druker, L. Hartwell, and R. Smith. 2005. Perspective: a program to improve protein biomarker discovery for cancer. *J. Proteome Res.* **4:** 1104–1109.

45. Steel, L. F., B. B. Haab, and S. M. Hanash. 2005. Methods of comparative proteomic profiling for disease diagnostics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **815:** 275–284.

46. Weston, A. D., and L. Hood. 2004. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* **3:** 179–196.

47. Desiere, F., E. W. Deutsch, A. I. Nesvizhskii, P. Mallick, N. L. King, J. K. Eng, A. Aderem, R. Boyle, E. Brunner, S. Donohoe, et al. 2005. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* **6:** R9.

48. Gilbert, K., S. Figueredo, X. Y. Meng, C. Yip, and E. T. Fung. 2004. Serum protein-expression profiling using the ProteinChip biomarker system. *Methods Mol. Biol.* **264:** 259–269.

49. Zhang, X., S. M. Leung, C. R. Morris, and M. K. Shigenaga. 2004. Evaluation of a novel, integrated approach using functionalized magnetic beads, bench-top MALDI-TOF-MS with prestructured sample supports, and pattern recognition software for profiling potential biomarkers in human plasma. *J. Biomol. Tech.* **15:** 167–175.

50. Zhang, H., E. C. Yi, X. J. Li, P. Mallick, K. S. Kelly-Spratt, C. D. Masselon, D. G. Camp, 2nd, R. D. Smith, C. J. Kemp, and R. Aebersold. 2005. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics.* **4:** 144–155.

51. Yook, S. H., Z. N. Oltvai, and A. L. Barabasi. 2004. Functional and topological characterization of protein interaction networks. *Proteomics.* **4:** 928–942.

52. Yu, H., Y. Xia, V. Trifonov, and M. Gerstein. 2006. Design principles of molecular networks revealed by global comparisons and composite motifs. *Genome Biol.* **7:** R55.

52a. Weiss, J. N., L. Yang, and Z. Qu. 2006. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Network perspectives of cardiovascular metablolism. *J. Lipid Res.* **47:** 2355–2366.

53. Ping, P. 2003. Identification of novel signaling complexes by functional proteomics. *Circ. Res.* **93:** 595–603.

54. Zhang, J., C. P. Baines, C. Zong, E. M. Cardwell, G. Wang, T. M. Vondriska, and P. Ping. 2005. Functional proteomic analysis of a three-tier PKCepsilon-Akt-eNOS signaling module in cardiac protection. *Am. J. Physiol. Heart Circ. Physiol.* **288:** H954–H961.

55. Anderson, L. 2005. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* **563:** 23–60.

56. Berhane, B. T., C. Zong, D. A. Liem, A. Huang, S. Le, R. D. Edmondson, R. C. Jones, X. Qiao, J. P. Whitelegge, P. Ping, et al. 2005. Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project pilot phase. *Proteomics.* **5:** 3520–3530.

57. Ganesh, S. K., K. A. Skelding, L. Mehta, K. O'Neill, J. Joo, G. Zheng, J. Goldstein, R. Simari, E. Billings, N. L. Geller, et al. 2004. Rationale and study design of the CardioGene Study: genomics of in-stent restenosis. *Pharmacogenomics.* **5:** 952–1004.

58. Coveney, P. V., and P. W. Fowler. 2005. Modelling biological complexity: a physical scientist's perspective. *J. R. Soc. Interface.* **2:** 267–280.

59. Ideker, T., and D. Lauffenburger. 2003. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* **21:** 255–262.

60. Joyce, A. R., and B. O. Palsson. 2006. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol.* **7:** 198–210.

61. Ma'ayan, A., R. D. Blitzer, and R. Iyengar. 2005. Toward predictive models of mammalian cells. *Annu. Rev. Biophys. Biomol. Struct.* **34:** 319–349.

62. Cox, B., T. Kislinger, and A. Emili. 2005. Integrating gene and protein expression data: pattern analysis and profile mining. *Methods.* **35:** 303–314.

63. Greenbaum, D., R. Jansen, and M. Gerstein. 2002. Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics.* **18:** 585–596.

64. Griffin, T. J., S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood, and R. Aebersold. 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae.* *Mol. Cell. Proteomics.* **1:** 323–333.

65. Kasukawa, T., S. Katayama, H. Kawaji, H. Suzuki, D. A. Hume, and Y. Hayashizaki. 2004. Construction of representative transcript and protein sets of human, mouse, and rat as a platform for their transcriptome and proteome analysis. *Genomics.* **84:** 913–921.

66. Maziarz, M., C. Chung, D. J. Drucker, and A. Emili. 2005. Integrating global proteomic and genomic expression profiles generated from islet alpha cells: opportunities and challenges to deriving reliable biological inferences. *Mol. Cell. Proteomics.* **4:** 458–474.

67. Tian, Q., S. B. Stepaniants, M. Mao, L. Weng, M. C. Feetham, M. J. Doyle, E. C. Yi, H. Dai, V. Thorsson, J. Eng, et al. 2004. Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol. Cell. Proteomics.* **3:** 960–969.

68. Washburn, M. P., A. Koller, G. Oshiro, R. R. Ulaszek, D. Plouffe, C. Deciu, E. Winzeler, and J. R. Yates, 3rd. 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae.* *Proc. Natl. Acad. Sci. USA.* **100:** 3107–3112.

69. Whitehead, K., A. Kish, M. Pan, A. Kaur, D. J. Reiss, N. King, L. Hohmann, J. DiRuggiero, and N. S. Baliga. 2006. An integrated systems approach for understanding cellular responses to gamma radiation. *Mol. Syst. Biol.* **2:** 47.

70. Jamshidi, N., and B. O. Palsson. 2006. Systems biology of SNPs. *Mol. Syst. Biol.* **2:** 38.