Status and Prospects for Discovery and Verification of New Biomarkers of Cardiovascular Disease by Proteomics

Robert E. Gerszten, Aarti Asnani, Steven A. Carr

Abstract: Despite unmet needs for cardiovascular biomarkers, few new protein markers have been approved by the US Food and Drug Administration for the diagnosis or screening of cardiovascular diseases. Mass spectrometry–based proteomics technologies are capable of identifying hundreds to thousands of proteins in cells, tissues, and biofluids. Proteomics may therefore provide the opportunity to elucidate new biomarkers and pathways without a prior known association with cardiovascular disease; however, important obstacles remain. In this review, we focus on emerging techniques that may form a coherently integrated pipeline to overcome present limitations to both the discovery and validation processes. (Circ Res. 2011;109:463-474.)

Key Words: proteomics ■ biomarkers ■ mass spectrometry

Current Status of Cardiovascular Biomarkers

Clinical biomarkers can serve a variety of functions that correspond to different stages in disease evolution. Biomarkers can assist in the care of patients without apparent disease (screening biomarkers), with suspected disease (diagnostic biomarkers), or with overt disease (prognostic biomarkers). Recent plasma or serum biomarkers that have been incorporated into cardiology practice successfully fall mostly into the category of diagnostic biomarkers, including troponin I and troponin T for myocardial injury. The diagnosis of acute coronary syndromes (ACS) currently relies on the measurement of circulating biomarkers of myocardial necrosis. As myocardium-specific structural proteins, cardiac troponins not only form the cornerstone for the diagnosis of myocardial infarction but also correlate with risk of mortality in non–ST-
elevation ACS. As a diagnostic tool, however, measurement of troponin is limited by the fact that peripheral blood levels may remain undetectable for several hours after the onset of myocardial injury. Several observational studies have confirmed that only 20% to 50% of patients with ACS have a positive troponin T level on presentation to the emergency department. Although the recent development of high-sensitivity troponins appears promising, the clinical implications of low-level values have yet to be defined and may result in decreased biomarker specificity for myocardial injury.

In addition to providing diagnostic information, determination of troponin levels can assist clinicians in tailoring therapeutic interventions for myocardial ischemia and infarction, particularly in the setting of non–ST-elevation ACS. Treatment with the glycoprotein IIb/IIIa inhibitor abciximab, for instance, was associated with reduced cardiovascular events and death only in those patients with ACS and an elevated troponin level. Troponin elevation has also been used to identify those patients with ACS who benefit from early cardiac catheterization. To that end, American College of Cardiology/American Heart Association guidelines for the management of patients with non–ST-elevation ACS recommend routine measurement of troponin to guide therapy.

Similarly, measurement of the biomarker B-type natriuretic peptide (BNP), a hormone released from the ventricles during myocardial stress, has become integral to the diagnosis and management of acute congestive heart failure. In a study of nearly 1600 patients presenting to the emergency department with acute dyspnea, an elevated BNP level (≥150 pg/mL) diagnosed congestive heart failure accurately in 83% of patients, whereas a low BNP level (≤50 pg/mL) effectively excluded congestive heart failure in 96% of patients. The high sensitivity and specificity of BNP are especially valuable in patients whose clinical examination findings may be unreliable. In fact, measurement of NT-proBNP (the N-terminal of the prohormone of BNP) in the emergency department has been shown to reduce rates of rehospitalization and overall resource utilization in patients with suspected congestive heart failure.

Limitations of Existing Biomarkers

Although biomarkers such as troponin are helpful in the diagnosis and management of irreversible myocardial injury, we currently have no satisfactory diagnostic markers of reversible myocardial ischemia, ie, either stable or unstable angina. Given the transient nature of electrocardiographic changes and the subjective nature of history taking, physicians will often rely on a stress test to confirm or exclude the diagnosis of myocardial ischemia; however, a standard exercise stress test has a sensitivity of only 60% (and less than 50% for single-vessel disease) and a specificity of only 70%. Biomarkers of ischemia could thus be used both for diagnosis and to monitor the efficacy of subsequent therapeutic interventions.

Furthermore, there are currently no widely accepted biomarkers for cardiovascular screening. This has become an active area of investigation, because the prevention of events in those at risk for cardiovascular disease (primary prevention) is likely to have a substantial impact on the overall public health burden. In the last decade, a number of circulating biomarkers have created enthusiasm, not only because of their success in predicting future cardiovascular events in ambulatory populations but also because of their mechanistic involvement in atherosclerosis-associated pathways. These include biomarkers associated with inflammation (C-reactive protein, interleukin-6, lipoprotein-associated phospholipase 2), hemostasis/thrombosis (fibrinogen, plasminogen activator inhibitor-1), neurohormonal activation (renin, BNP), insulin resistance (insulin, hemoglobin A1C), and endothelial dysfunction (homocysteine, urinary microalbuminuria). Whether these new biomarkers add information on top of existing risk factors or biomarkers is unclear, however, and thus, there is a pressing need to integrate proteomics techniques into biomarker discovery.

Proteomics Approaches to Cardiovascular Biomarker Discovery

Despite the limitations of cardiovascular biomarkers in use today, few new protein markers have been approved by the US Food and Drug Administration for the diagnosis and risk stratification of ACS or for cardiovascular screening. A similar situation exists for many other complex diseases, including cancer. Mass spectrometry (MS)–based proteomics technologies operated in the “discovery” mode (ie, untargeted, where the instrument performs data-dependent acquisition) are capable of identifying hundreds to thousands of proteins in cells, tissues, and biofluids (see MS-Based Biomarker Discovery, below). Discovery proteomics may therefore provide the opportunity to elucidate new biomarkers and pathways without a prior known association with cardiovascular disease. However, important obstacles must be considered in any discussion of how proteomics will facilitate biomarker discovery. First, analytic barriers exist in working with extremely complex mixtures such as human plasma. Like all of the “-omics” technologies that survey hundreds or thousands of signals or analytes in relatively small numbers of samples, many of the candidate biomarkers observed by proteomic methods are false discoveries. The term “false discovery” does not necessarily convey that the detection of differential abundance by proteomics is in error. Rather, many of the differences in protein abundance detected by proteomics may arise from interindividual variation in protein abundance and not from the underlying disease process under investigation. To identify which of the candidate biomarker
proteins are likely to be disease-relevant, it is essential that we develop robust methods to test large numbers of biomarker candidates emerging from discovery “-omics” studies using specific and quantitative measurements in relatively large patient cohorts for initial verification (eg, hundreds). As we detail below, new technologies are emerging that have great potential to overcome each of the aforementioned barriers.14–18

The plasma proteome is unique in that it does not represent a particular cellular genome but instead reflects the collective expression of all cellular genomes. It has thus far been poorly characterized. Three factors are responsible for the difficulty in fully characterizing the plasma proteome by MS. First, there is a dominance of a few high-abundance proteins in blood. A single protein, albumin, constitutes over 50% of the total protein mass and is present at approximately 35 to 60 mg/mL in humans.19 The top 22 most abundant proteins, including albumin and the immunoglobulins, constitute approximately 99% of the plasma proteome mass.20 A second major hurdle is the vast number of proteins and modified forms of these proteins that exist in blood. Estimates of the number of proteins in blood vary widely from 10 000 unique proteins to 1 000 000 proteins depending on whether the estimate attempts to take into account the number of variants due to proteolytic processing, posttranslational modifications, single-nucleotide polymorphisms, and splice variants that may exist. Another important impediment to characterizing the human plasma proteome is the very wide dynamic range in concentrations over which these proteins are found, spanning an estimated 11 orders of magnitude, from >600 μmol/L to concentrations in the range of low femtomoles per liter of blood.19 Many of the biologically interesting molecules relevant to cardiovascular disease are low-abundance proteins. For example, cardiac markers such as the troponins are found in the nanomolar range and tumor necrosis factor-α in the femtomolar range even when elevated in pathological states. Many lower-abundance proteins in plasma appear to be intracellular or membrane proteins, present as a result of cellular signaling, tissue disruption, remodeling, apoptosis, or necrosis.

MS-Based Biomarker Discovery

To understand the impact of these factors on the results that can be obtained from proteomics analyses of clinical samples, it is necessary to briefly describe current state-of-the-art proteomics experiments and to provide some sense of their capabilities and their limitations for biomarker discovery.21 Although many MS methods have been used in all areas of disease biomarker discovery,14 here we focus predominantly on liquid-chromatography (LC) tandem MS (LC-MS/MS). LC-MS/MS, especially when combined with an additional chromatographic step of peptide or protein fractionation before the final online LC-MS/MS analysis (so-called multidimensional LC-MS/MS), is currently the only technology that has been demonstrated to robustly detect and identify tens of thousands of peptides and thousands of proteins in tissue, proximal fluids, and plasma samples.22,23 Sensitivity and relative comprehensiveness of peptide/protein identification are of central importance in biomarker discovery studies, because proteins specifically related to the disease mechanism are presumed to be present at low levels, particularly in proximal fluids and most especially in peripheral blood. Sample ionization is usually accomplished by electrospray, which is ideally suited for online LC-MS/MS analysis. The reproducibility and robustness of LC-MS/MS methods have also been evaluated carefully, and both interlaboratory and intralaboratory and metrics to assess performance have been established.24,25 Matrix-assisted laser desorption/ionization (MALDI) MS can also be mated to multidimensional separations, albeit with some significant tradeoffs.26 Although direct connection of separation devices to MALDI has been attempted, performance approaching that of online LC-MS/MS requires that peptides be fractionated offline, spotted on MALDI targets, and then introduced into the mass spectrometer for analysis, a significant disadvantage relative to online LC-MS/MS. Two-dimensional gel electrophoresis followed by MS analysis of differentially abundant gel spots is also a multidimensional separation approach that has been widely used in biomarker discovery,27 but typically, it can identify only a few hundred proteins in biological samples.

Differential peptide/protein expression analysis by LC-MS/MS involves the comparison of case samples to control (or comparator) samples (Figure 1). Proteins in the samples are reduced and alkylated to cleave disulfide bonds and block the cysteine residues, after which the proteins are cleaved to peptides with the enzyme trypsin. Peptides in the resulting mixture are temporally separated with a reversed-phase chromatographic column connected directly to the mass spectrometer. In the mass spectrometer, peptides eluting from the column are ionized by electrospray and analyzed on a high-performance MS system to provide the mass and charge (m/z) of the intact peptides. Intact peptide ions are also fragmented inside the MS system through physical interaction with a gas to produce sequence-informative fragment ions. The volume of data generated during the course of a single LC-MS/MS analysis is enormous, typically producing 4000 MS scans and 30 000 MS/MS scans. The widespread availability of robust software packages capable of reliably analyzing the resulting MS and MS/MS data has been critical to the expansion of proteomics techniques, because it is no longer possible for an analyst to interrogate all of the data manually.

The MS instrument scan cycle in a typical instrument starts with the acquisition of a full-scan mass spectrum in a period of approximately 1 second (Figure 2). The mass spectrum is a record of the mass-to-charge ratio (m/z) and intensities of the ions observed during that 1-second period of the chromatographic separation. This sampling is performed repetitively over the entire course of the LC-MS/MS analysis, which generally lasts between 60 and 180 minutes. The more complex the sample, the longer the separation time used. Importantly, when multidimensional separations have been used, each fraction from the first dimensional separation requires 60 to 180 minutes of LC-MS/MS time to analyze. An example of a mass spectrum is shown in Figure 2 recorded at 70.55 minutes during a 90-minute gradient separation of a digested representative tissue lysate.

The selection of which ions to fragment for sequencing is done by one of the instrument’s onboard processors, without
human intervention, in what has become known as a “data-dependent” experiment. The onboard processor identifies the peptide mass and charge in the mass spectrum and selects the top $n$ of these ions to fragment; $n$ is defined by the user and ranges from 3 to 20 depending on the data acquisition speed and the sensitivity of the specific instrument used. In the example shown in Figure 2, the top 8 most abundant ions in this spectrum were automatically selected and fragmented for sequencing in approximately 2 seconds, each time creating additional spectra. After acquisition of the current block of $n$ MS/MS spectra, the scan cycle is repeated continuously for the duration of the online LC separation, each time starting with acquisition of a full-scan MS spectrum followed by $n$ more MS/MS spectra. These data subsequently are analyzed by the analysis software (eg, Mascot, Spectrum Mill, Sequest, XTandem).

**Decreasing Sample Complexity Is Essential to Gaining Depth of Proteome Detection**

Despite tremendous advances in MS technology over the past several years, the number of detectable mass and charge ratios in each full-scan mass spectrum acquired during analysis of highly complex biological samples (such as biofluids or tissues that are used for biomarker discovery) routinely exceeds the value of $n$. In addition, not all MS/MS spectra yield an interpretable sequence. Multiple technical reasons contribute to this phenomenon, and the consequences are important: 50% or more of the mass-to-charge ratio peaks in an LC-MS/MS analysis can go unsequenced or uninterpreted by the analysis software. The more complex the sample, the more peptides that will elute at any given moment in time from the chromatographic column, and the greater the potential to lose information.

To deal with this problem, investigators use additional stages of separation ahead of the ultimate online reverse-phase LC separation to reduce the complexity of the sample presented to the LC-MS/MS instrument. A number of useful separation methods are orthogonal, or distinct, from the final reverse-phase LC separation. Typically, these techniques are performed at the peptide level with strong cation exchange,28
Figure 2. The LC-MS/MS data acquisition process commonly used in biomarker discovery. **Top,** Total ion current trace recorded during online LC-MS/MS analysis of a tissue lysate. The total ion current is similar to a UV trace from the chromatograph, but instead of a plot of absorbance versus time, the total ion current is a plot of the sum of the ions entering the MS versus time. **Middle,** Mass spectrum (plot of mass-to-charge ratio \[m/z\] versus time) recorded in approximately 1 second at time \(t = 70.55\) minutes during the LC-MS experiment. The peaks correspond to the mass-to-charge ratio of peptides eluting from the column at this moment in time. On modern high-performance instruments, the MS data are recorded at high resolution and with high mass precision, typically 30 000 to 60 000 resolution and 2 to 5 parts per million mass accuracy. Both high resolution and mass accuracy are important in the analysis of highly complex mixtures such as digested plasma. In such complex mixtures, many peptides of different amino acid sequence will have the same nominal molecular mass (ie, within 1 mass unit) but will differ in their accurate molecular masses. High instrument resolving power enables peptides of similar but nonidentical mass to be separated on the mass-to-charge ratio scale, and high mass accuracy facilitates their identification by the software by narrowing the list of possible amino acid compositions that could correspond to the observed accurate mass. **Bottom,** One of 8 MS/MS spectra recorded in approximately 100 ms by the onboard processors in the MS system. In a typical experiment, the top 8 to 10 most abundant ions in the current mass spectrum are sequentially mass selected and their fragmentation spectra recorded for sequencing automatically during a 1- to 3-second period. After acquisition of the current block of \(n\) MS/MS spectra, the scan cycle is repeated continuously for the duration of the online LC separation, each time starting with acquisition of a full-scan MS spectrum followed by \(n\) more MS/MS spectra. These data are subsequently analyzed by the analysis software (eg, Mascot, Spectrum Mill, Sequest, XTandem). The peptide sequence determined by the data analysis software is shown. The \(b\) and \(y\) ions correspond to fragmentation along the peptide backbone at amide bonds with charge retained on the N-terminal fragment or C-terminal fragment, respectively.
high pH reverse phase,\textsuperscript{29} or off-gel electrophoresis,\textsuperscript{30} but protein-level separation techniques are also used, including 1-dimensional gel separations coupled with digestion of the entire lane cut into 10 to 20 bands. However, even in highly fractionated samples, many peptides are not subjected to sequencing by MS/MS, as noted previously. Although these types of multidimensional separation techniques reduce the complexity of the sample for subsequent LC-MS/MS analysis and allow for broader and more sensitive analyses, they in turn invoke a significant additional practical limitation, because this process may generate up to 100 subfractions from a given starting sample. Thus, a single sample can occupy an expensive LC-MS/MS instrument for 1 to 2 weeks. Given the sensitivity and throughput of presently available techniques, proteomics discovery studies are most useful and practical when applied to relatively small (eg, tens) of well-phenotyped samples. Ion mobility spectrometry is showing promise to improve analysis of complex peptide samples using more limited sample fractionation before MS.\textsuperscript{31} In ion mobility spectrometry, peptide ions are separated in the gas phase on a millisecond time scale on the basis of their charge, size, and shape. Ion mobility spectrometry, therefore, provides another dimension of peptide separation without an additional chromatographic step. Routine use of ion mobility spectrometry in proteomics awaits widespread implementation of the necessary hardware and software by MS vendors.

Another important consideration in the application of an LC-MS/MS–based experimental approach is that the sample the MS system is evaluating at any given time is not static but changing dynamically during the analysis. Small changes in the elution order or biochemical background observed at any given time in the chromatography alter what the MS instrument observes to be the top \( n \) peptides selected for sequencing in a data-dependent experiment. Furthermore, the number of peptides that can be selected in a top \( n \) experiment is limited by the speed and sensitivity of the specific MS system used, as well as by the chromatographic peak width. As a result, even in highly fractionated samples, many peptides are not subjected to sequencing by MS/MS, and peptides that are sequenced in 1 sample are not sequenced in other samples despite being present. This is quite different from microarray expression profiling, for example, in which lack of detection of a signal more definitively suggests that the transcript is absent. In the case of LC-MS/MS data, lack of a signal can either mean that the peptide is not present or that it was not sequenced by MS/MS. Fortunately, detection of a protein (versus a peptide from that protein) does not require detection of exactly the same peptides across samples. Proteins typically yield multiple peptides on digestion, and some subset of these are detected in individual runs. These peptides allow detection and relative quantification of the proteins across samples. The greater the number of peptides identified from each protein, the greater the confidence in both the identification and relative quantification.

The Value of Pattern Recognition in Proteomics

Approaches have been developed to try to extract additional information from LC-MS experiments to overcome this inherent sampling limitation. In these methods, patterns of mass-to-charge ratio obtained at high mass accuracy and at high resolution in the MS scans, together with the ion signal intensities and retention times, are recorded for each sample. Mathematical approaches are then used, after data acquisition, to align these patterns across the LC-MS analyses.\textsuperscript{32} Peptides observed to change in abundance from 1 sample to another are searched for and identified in separate MS/MS analyses that have been performed on these same samples. We have combined this “accurate mass and time” pattern approach with simultaneous acquisition of MS/MS data on the same instrument for peptides scattered throughout the chromatographic separation. These “landmark” peptides enable more accurate peak alignment across large samples and identification of differential signals of possible interest as biomarkers.\textsuperscript{33} Peptides of interest for which MS/MS data were not acquired as landmarks are subsequently analyzed by targeted LC-MS/MS by use of inclusion lists of accurate mass and charge.\textsuperscript{34} Adoption of high-resolution, accurate mass pattern–based approaches has been relatively slow and is largely limited at present by the lack of easy-to-use software that is instrument-vendor neutral, as well as the difficulty in handling data from highly fractionated samples.

Reducing Protein Dynamic Range Through Abundant Protein Depletion

The choice of sample type (eg, plasma, tissue, proximal fluids, perfusates) and the experimental design naturally have a considerable impact on achievable results. As alluded to above, the presence of a small number of highly abundant proteins in plasma or serum is a major complicating issue for proteomic analysis. Peptides from these highly abundant proteins are observed in essentially every fraction, regardless of separation modality, and the intensities of the peptide ions often dominate the spectra, which results in the sequencing of highly abundant but largely irrelevant proteins. To reduce the impact of these proteins on the breadth and depth of the plasma proteome detected, abundant plasma proteins are extracted by immunoaffinity depletion columns that remove the top 14 or more proteins.\textsuperscript{35–38} Even with the use of these columns, the number of proteins confidently detected in plasma (identified with 2 or more peptides) by a multidimensional separation approach that combines strong cation exchange (80 fractions) followed by LC-MS/MS analysis of each fraction is approximately 1000 to 1200 unique proteins, excluding immunoglobulins. By contrast, tissue or tissue surrogates (tissue interstitial fluid) may yield as many as 4000 to >6000 confidently identified proteins when comparable methodologies are used (unpublished data, Carr laboratory). This suggests that there are additional tiers of relatively highly abundant proteins in blood below the 20 or so most abundant constituents, and it has been hypothesized that that the removal of perhaps the top 100 proteins will be necessary to achieve greater depth of protein detection by discovery methods in plasma.\textsuperscript{38} Other approaches that have been reported are the use of peptide-coated beads to absorb abundant proteins and thereby compress the dynamic range of the sample. These approaches have promise, but their reproducibility and robustness still need to be demonstrated. A concern regarding the use of protein-depletion technologies is
the unintended removal of proteins of possible interest through binding (specific or nonspecific) to the proteins targeted for depletion or as a result of losses after sample workup after depletion. Our studies confirm that nonspecific losses do occur when these columns are used\(^{17,18}\); however, these studies also demonstrate that the losses are offset by the 10 to 20-fold reduction in sample complexity. Commercially available depletion columns are highly reproducible and can be used effectively to process large numbers of patient samples.

**Emerging Proteomics Techniques for Candidate Biomarker Verification and Validation**

Given the hundreds of potential biomarkers that arise from discovery “-omics” experiments, biomarker verification is essential to identify those few candidate proteins that merit traditional validation studies on clinically approved analysis platforms.\(^{15,39}\) Until recently, verification technologies capable of testing large numbers of protein biomarker candidates through a targeted approach have not been available. This deficiency has contributed significantly to the lack of translation of so-called biomarker discoveries into the clinic. Without appropriate credentialing, no effort can or should be expected to advance these early discoveries into expensive, time-consuming clinical validation studies.\(^{14}\) Verification and validation of putative biomarkers with antibody-based techniques is, of course, already common practice; however, antibody pairs (eg, for sandwich immunoassays) of sufficient specificity often exist for only a limited number of candidate proteins, which highlights the need for non–antibody-based verification technologies. Furthermore, a typical immunoassay may require 100 \(\mu\)L or more of plasma or serum. Although multiplexed antibody-based systems gained initial enthusiasm, there has been an increasing concern with regard to interferences and specificity when assays are performed with pooled antibodies. Ideally, new techniques must have reasonable assay development timelines and low assay development cost, be effectively multiplexed to assess tens of proteins in a single analysis, require minimal volumes of patient sample, and achieve a throughput of hundreds of patient plasma samples with good assay precision.

In an effort to respond to the existing throughput limitations of protein biomarker verification, there have been several key advances in MS-based verification technologies. Stable isotope dilution–multiple reaction monitoring (MRM) MS has emerged as a valuable technique to perform quantitative assays of candidate biomarkers.\(^{17}\) This technology relies on measurement of peptides created by trypsin digestion of whole proteins. Two to 5 peptides are generally chosen to represent each protein biomarker for analysis.\(^{16–18}\) Peptides are usually selected because they have sequences unique to the protein of interest (“proteotypic peptide”) and are among the peptides with the highest response in the MS system to afford the greatest sensitivity (“signature peptide”).\(^{40}\) Synthetic isotope–labeled versions of each peptide are used as internal standards, which enables protein concentration to be measured by comparing the signals from the exogenous labeled and endogenous unlabeled species. This method has several advantages over conventional immunoassays, including structural specificity of analyte detection and the ability to measure dozens of proteins in a single analysis with <100 \(\mu\)L of plasma. A combination of abundant protein depletion and minimal fractionation of tryptic peptides by strong cation exchange followed by stable isotope dilution–MRM-MS provides limits of quantitation in the 1- to 20-ng/mL range.\(^{17,18}\) This technique has led to the development of assays for known markers of cardiovascular disease, such as cardiac troponin I, as well as novel biomarkers for which antibody reagents are not available, such as interleukin-33.\(^{41}\)

Although this approach has resulted in the configuration of new protein assays, the need for upfront sample processing and limitations to MS sensitivity remain barriers to biomarker verification. To address these limitations, stable isotope standards with capture by anti-peptide antibodies (SISCAPA) has emerged as a method of combining specific immunoaffinity enrichment of a target peptide with the structural specificity and quantitative capabilities of stable isotope dilution–MRM-MS.\(^{42}\) In this approach, anti-peptide antibodies are made against the selected tryptic peptides from protein biomarkers of interest (Figure 3). After digestion of plasma and the addition of known amounts of stable isotope–labeled standard peptide, both added and sample-derived versions are specifically enriched and the relative amounts measured by MRM. In this context, the mass spectrometer essentially serves as the second antibody. More than a thousand-fold enrichment can be achieved for plasma digest peptides with this approach,\(^{43}\) and SISCAPA assays can achieve limits of quantitation in plasma in the low nanograms per milliliter range with coefficients of variation less than 20%.\(^{41}\) In addition, throughput can be improved significantly by coupling SISCAPA to magnetic bead–handling robotic devices that automate peptide capture, wash, and elution steps. Using this approach, investigators have configured assays to cardiac troponin I and to interleukin-33, a potentially novel protein biomarker of cardiovascular disease (Figure 3).\(^{41}\)

At present, MRM-MS– and SISCAPA-based assays are best suited to the research environment in which tens to hundreds of candidate biomarker proteins are to be evaluated for the purposes of candidate verification.\(^{14}\) That said, MRM-MS and SISCAPA-MRM-MS assays are already making their way into the clinical laboratory. The first clinical-grade SISCAPA assay has already been constructed and is in use for measuring thyroglobulin.\(^{44}\) Liquid-handling robotic devices for plasma sample digestion and antibody-based peptide capture are also being implemented. A group led by the National Cancer Institute’s Clinical Proteomic Technologies in Cancer Program has explored the necessary analytic hurdles that must be passed to obtain Food and Drug Administration approval of multiplexed SISCAPA assays.\(^{45}\) SISCAPA assays are also being adopted by biotechnology and pharmacological industries to measure pharmacokinetic properties of proteins of interest.\(^{46}\) Analysis times are currently on the order of 15 to 60 minutes per sample with conventional LC systems, and shorter analysis times are possible with integrated LC-chip interfaces. The current cost relative to clinical immunoassays is high but is likely to drop as the technology is adopted and refined. The
The principal cost drivers are the additional reagents (including peptides, beads, and enzymes) needed for plasma sample processing. The antibodies are no more expensive to make than conventional antibodies, and the success rate for obtaining an antibody of sufficient quality for configuring a SISCAPA assay is higher than that for obtaining a reagent suitable for capture of the corresponding protein.\textsuperscript{47} The cost of the LC-MS/MS instrumentation is on par with the cost of high-end immunoassay systems. A tremendous advantage of the MRM-MS and SISCAPA-MRM-MS technologies over conventional immunoassay approaches is that they may be highly multiplexed without increasing the amount of patient sample required.

**Emerging Capture and Detection Platforms**

In addition to evolving MS-based methods, a number of emerging microarray techniques have sought to address the limitations of current proteomics technology. “Self-assembling” protein microarrays consist of cDNAs that are printed onto glass slides and subsequently translated to target proteins in situ by mammalian reticulocyte lysates.\textsuperscript{48} In these nucleic acid programmable protein arrays, proteins of interest are marked with epitope tags and immobilized on the array. This obviates the need for further purification steps, which allows for increased throughput. An antibody-based protein array that uses cDNAs preferentially expressed in ruptured human atherosclerotic plaques was recently used to search for autoantibodies in coronary atherosclerosis.\textsuperscript{49} Although 2 cDNA products were reported as potential biomarkers of early acute myocardial infarction, these antibodies appear to be directed against random peptides rather than self-antigens, and whether these antibodies exist in vivo remains unclear. Although nucleic acid–based protein arrays are a promising development in high-throughput proteomics, the ex vivo nature of such experiments presents a challenge for translating these findings into clinically useful biomarkers.

The sensitivity and specificity of high-throughput protein assays continue to limit the applicability of such proteomics technologies, particularly with regard to low-abundance proteins. One potential solution involves the use of an aptamer microarray in combination with a nanoporous sol-gel.\textsuperscript{50} Aptamers are DNA or RNA oligonucleotides that may bind...
specifically to proteins. Once a protein is captured on the sol-gel, it is digested in situ, and the digest is subsequently analyzed by MS. In a proof-of-principle study, target protein present at 0.001% the concentration of total serum protein was identified without pretreatment of the sample. Using the unique kinetic properties of aptamer binding, slow off-rate modified aptamers (SOMAmers) have been developed to further increase the specificity of such assays.51 These aptamers have a low dissociation constant for their target proteins but high dissociation constants for other proteins in plasma. Although significant advances have been made in aptamer technology, the specificity of each aptamer reagent for its target protein must be proven in direct analogy to what is required of antibodies. To date, only a few such demonstrations have been described. In a similar effort to improve the specificity of proteomics experiments, Reddy and Kodadek52 have developed microarrays comprising several thousand peptoids (peptide mimetics whose side chains are appended to the nitrogen atom of the peptide backbone rather than to the α-carbons, as they are in amino acids). Peptoids may bind to specific proteins with high affinity, which creates a unique molecular “fingerprint.”52 This approach can then be combined with an antibody-based assay to identify proteins in complex mixtures such as plasma.

Sample Selection

The analysis of proximal fluids is particularly advantageous for discovery samples. As their name implies, these fluids are near or in contact with the site of interest and are therefore likely to have elevated concentrations of proteins that are being actively shed or secreted as a result of a given (patho)physiological process. For example, nipple aspirates have been used to identify markers of breast malignancy.53 In the case of cardiovascular disease, we and others have studied coronary sinus blood as a source of enriched proteins and small molecules.54 Others have suggested the use of “synthetic” proximal fluids, created by isolated organ perfusion in animal models. In one example, the levels of heart-specific proteins were observed to increase in the perfusate of rat hearts subjected to repeated cycles of ischemia and reperfusion.55 In a similar model, the coronary effluent of isolated murine hearts was found to be enriched in cardiac myosin-binding protein C after ischemia, which suggests that this protein may serve as an early marker of cardiac injury.56 Several groups have used isolated organ perfusion to probe the mechanisms involved in ischemic preconditioning, specifically with regard to changes in the mitochondrial proteome.57–59 Importantly, the detectability of these proteins was greatly enhanced because of the much reduced levels of contaminating plasma proteins. After identification in the proximal fluid, or even the organ of interest, targeted LC-MS/MS techniques can then be used to assay for the proteins in circulating blood.

Recent Proteomics Studies in Cardiovascular Disease

To date, the application of proteomics to cardiovascular disease has largely focused on efforts to gain biological insight into disease mechanisms. To better understand coronary atherosclerosis, for example, investigators have evaluated the proteome of apolipoproteins associated with HDL. HDL-associated proteins not only participate in lipid metabolism, but they are also thought to have important functional roles in inflammation, immune system activation, and hemostasis. A “shotgun” approach to investigate the HDL proteome has been used in a small cohort of patients with known coronary artery disease.60 In these patients, HDL, was enriched in apolipoprotein E compared with control subjects, which confirms the notion that apolipoprotein E is essential to atherogenesis and could potentially be used as a marker for subclinical cardiovascular disease. Similarly, a proteomic analysis of atherosclerotic plaque from patients undergoing carotid endarterectomy demonstrated that specific isoforms of α1-antitrypsin may distinguish plaques that contain thrombus from those that are advanced but stable.61

Protein profiling experiments have also been performed to elucidate the mechanisms that lead to ventricular hypertrophy and congestive heart failure. In a rat model of hypertension-induced left ventricular hypertrophy, proteomic assessment of left ventricular tissue isolated from those rats with early left ventricular hypertrophy revealed 2 unique proteins, calscarcin-1 and ubiquinone biosynthesis protein COQ7, which were not found in left ventricular tissue from control rats.62 Interestingly, calscarcin-1 is a negative regulator of the calcineurin/NFAT (nuclear factor of activated T cells) pathway and may contribute to increased oxidative stress. With regard to right-sided heart failure, a recent proteomics study identified Fhl-1, a protein involved in muscle development, as an early marker of hypoxia-induced pulmonary hypertension.63 Although such studies are illuminating from a biological perspective, they remain limited to animal studies and small cohorts of patients and require validation in larger populations before they can be applied in a clinical setting.

Future Directions and Prospects

The majority of cardiovascular biomarkers in use today participate in well-established pathways associated with atherosclerosis, such as inflammation, hemostasis, and cholesterol metabolism; however, such biomarkers may only provide information that is correlated with what is already known or being measured and thus may not significantly increase predictive value. For example, a recent investigation from the Framingham Heart Study evaluated 10 cardiovascular biomarkers in more than 3000 people followed up for nearly a decade.64 Multiple biomarkers were found to be statistically significant predictors of death (C-reactive protein, BNP, urinary albumin excretion, renin, homocysteine) or cardiovascular events (BNP, urinary albumin excretion). When the biomarkers were combined into a multimarker score, individuals with high multimarker scores had a 4-fold increased risk of death and a 2-fold increased risk of major cardiovascular events compared with persons with low multimarker scores. However, the multimarker score was associated with only a modest increase in the area under the receiver operating characteristic curve compared with a risk score based on conventional risk factors alone. Thus, the predictive information conveyed by the biomarkers was already
being captured by the clinical risk factors alone; that is, they were interrogating similar pathways and were highly correlated. Moving forward, it is anticipated that proteomics may generate multiple novel biomarkers along new pathways that would improve our ability to diagnose more subtle manifestations of cardiovascular disease.

Such a multimarker approach in which distinct pathobiological axes are assessed biochemically may provide the optimal diagnostic tool to identify early ACS. Given that troponin is a marker of myocyte necrosis, C-reactive protein a marker of inflammation, and BNP a marker of left ventricular overload, it was hypothesized that simultaneous assessment of all 3 markers would offer complementary prognostic information and enable clinicians to risk stratify patients with ACS more effectively. When patients presenting with ACS were categorized on the basis of the number of elevated biomarkers at presentation, there was a near doubling of the mortality risk for each additional biomarker that was elevated, and this finding was subsequently validated in an independent cohort.65

Other biologically relevant biomarkers have also been successfully incorporated into a multimarker approach, notably myeloperoxidase (MPO), a protein expressed by neutrophils during immune system activation.66 A multimarker that consists of troponin, C-reactive protein, MPO, and soluble CD40 ligand accurately predicted 6-month cardiovascular event and mortality rates. Interestingly, MPO levels did not correlate with the other 3 markers or with ECG changes, and levels were found to be elevated in those patients at risk for cardiovascular events who had a normal troponin level on presentation. In this fashion, emerging proteomics techniques will allow us to “overlap” new biomarkers on existing multimarker scores, thus providing additional information for cardiovascular disease diagnosis and management.67

Sources of Funding
The authors gratefully acknowledge support from the National Institutes of Health, National Heart, Lung, and Blood Institute’s Proteomics Center Contract (HHSN268201000033C) and R01HL096738-02.

Disclosures
None.

References


Status and Prospects for Discovery and Verification of New Biomarkers of Cardiovascular Disease by Proteomics

Robert E. Gerszten, Aarti Asnani and Steven A. Carr

doi: 10.1161/CIRCRESAHA.110.225003

_Circulation Research_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/109/4/463

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation Research_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation Research_ is online at:
http://circres.ahajournals.org//subscriptions/