Editorial

Substrate-Guided Proteomics Enhances Degradome Resolution

D. Kent Arrell, PhD; Andre Terzic, MD, PhD

High-throughput technologies offer unrecognized insights into the genesis of human disease, propelling innovation across a spectrum of biomedical specialties. The evolution of molecular diagnostics continues to shape deployment of modern therapies, including identification of markers of disease prognosis, predictors of therapeutic response, and determinants of optimal clinical management, all recognized cornerstones of modern medical practice. In the postgenomic era, proteomic paradigms have demonstrated particular utility in advancing the practice of personalized disease prediction, diagnosis, and therapy, with systems-based approaches facilitating functional deconvolution of proteomes despite their inherent biological complexity. The ever-expanding proteomic toolkit encompasses an array of sophisticated technologies with the flexibility to interrogate a totality of putative protein targets, thereby enabling systematic resolution of proteome structures and functions. Indeed, mapping and decoding of proteome landscapes in health and disease provide the foundation to iteratively collate, integrate, and prioritize large-scale raw data. Thus, high-throughput proteomics offers a robust, adaptable technological foundation from which to systematically comprehend the underpinnings of (patho)physiological processes.

Innovative proteomic methods are designed to separate and identify proteins with high fidelity and to assess protein abundance, structure, posttranslational modifications, and interactions, enabling charting of the global proteome (ie, the protein complement of a genome; Figure). Progress in proteomic research has benefited from the application of stringent criteria in assigning peptide and protein identities for comprehensive proteome cartography, such that perturbed or modified subproteomes within reconstituted protein networks can now be extracted at unprecedented rates for interpretation of their biological relevance. Expansion of proteomic studies has also been propelled by the concomitant development of high-resolution, high-mass-accuracy instrumentation, together with next-generation bioinformatics platforms for data analysis and systems interpretation. Existing approaches, however, may fall short of offering adequate experimental solutions arising from particular protein physicochemical constraints, limiting the otherwise far-reaching potential of proteomic science.

Application of a substrate-guided proteomics strategy to identify tissue-specific matrix metalloproteinase (MMP) targets, presented in this issue of Circulation: Cardiovascular Genetics by Stegemann et al, offers insight into circumventing such constraints. To minimize peptide combinatorial complexity for bioinformatic interpretation while maximizing protein coverage by yielding a high number of peptides measurable within instrumental precursor ion mass:charge ratio constraints, most peptide-based proteomic studies rely on digestion with a single protease of known amino acid cleavage specificity, most often trypsin, or a combination of proteases with shared specificity to increase cleavage efficiency such as trypsin and endoprotease LysC. This study, in contrast, leverages a technique analogous to one applied by the Overall laboratory to identify MMP-2 substrates in a mouse fibroblast secretome using endoprotease GluC in combination with MMP-2 cleavage. Stegemann et al investigate the vascular effects of MMP-3, MMP-9, and MMP-14, representing stromelysins, gelatinases, and collagenases, respectively, on human radial arteries, and of differing concentrations of MMP-9 on human ascending aorta. To facilitate peptide detection and identification, a sequential protease schedule was initiated by exploiting the natural substrate repertoire, or degradome, to yield isoform-specific MMP cleavage sites, followed by an enzyme with a different, well-characterized amino acid specificity, in this case trypsin. In this way, the diversity of degradome portfolio-containing MMP cleavage sites was contained within smaller trypsic peptide fragments more amenable to mass spectrometric analysis and bioinformatic interpretation. Indeed, in the study highlighted here, in vitro cleavage of fibronectin with MMP-3 alone resulted in a multitude of high-mass fragments that would preclude their detection by standard peptide-based mass spectrometry approaches. Further confounding matters, a lack of MMP peptide-bond hydrolysis sequence specificity leads to combinatorial effects on bioinformatic search parameters, which are computationally prohibitive to apply directly to detection of MMP-specific cleavage. By combining the natural degradome with tryptic digestion, however, search parameters were simplified by specifically screening for and detecting semitryptic peptides, in this instance, the result of MMP cleavage at one terminus and trypsin cleavage at the other, while ignoring peptides possessing fully trypsic termini. Applied to MMP substrates of the human vasculature, this...
to tissue healing, regeneration, and even cell transdifferentiation. In the vasculature, MMPs influence proliferation, migration, and apoptosis of smooth muscle, endothelial, and inflammatory cells, and thus play a homeostatic role in angiogenesis and thrombus resolution. Furthermore, MMP activity is implicated in vascular diseases, including atherosclerosis, aortic aneurysm, plaque rupture, and neointimal hyperplasia. Knowledge of the extent and specificity of MMP substrates in human vasculature may therefore provide insights applicable to a multiplicity of vascular interstitial processes.

Detecting semitryptic peptides from a sequential MMP/trypsin protease schedule, previously unavailable evidence of the MMP degradome was acquired, with a total of 74 distinct nontryptic cleavage sites identified by mass spectrometry. A summary of these sites indicates that 68 were cleaved by MMP-3, 13 by MMP-9, and 12 by MMP-14, with 9 sites shared between MMP-3 and MMP-9, 8 in common between MMP-3 and MMP-14, and 1 recognized and cleaved by all 3 MMP isoforms. This expansion of the breadth of MMP-specific ECM cleavage sites enriches current understanding of the MMP substrate spectrum in human vasculature. Beyond validating previously identified MMP substrate protein cleavage sites, several novel sites and substrates were identified, such as emelin-1 and tenascin-X, while particular proteins known to be substrates of 1 MMP isoform were also found to be susceptible to cleavage by other MMPs. Some substrates were digested at previously unidentified amino acid sites, whereas others exhibited endoprotease cleavage specificity identical to that of other MMP isoforms. For instance, an established MMP-7 cleavage site generating endostatin from collagen α-1 (XVIII) was replicated in both MMP-3 and MMP-14 experiments. Moreover, the first evidence for tenascin-C and periostrin cleavage by both MMP-3 and MMP-14 was presented for human arterial tissue. A comprehensive catalog was thereby assembled from proteomic cartography of MMP-dependent human vascular ECM cleavage sites.

Inception of an endoprotease recognition site database provides a resource for interrogation of putative actionable targets in the context of MMP-dependent vascular proteome remodeling. A critical step in maximizing database utility is fidelity in matching cleavage sites with respective MMP isoforms. To this end, strategies based on divalent cation-dependent and isoform-specific pharmacological modulation of metalloproteinase catalytic activity may streamline the precision of reported cleavage site specificity, minimizing or avoiding potential confounding influences arising from known MMP cascade effects (eg, MMP-3 proteolytic activation of MMP-9). Assignment of unique site-specific signatures would ensure accuracy in ascribing activity of distinct MMP isoforms to specific biological processes, an important consideration in view of their considerable substrate overlap, which will further increase as data for additional MMP isoforms are incorporated. Documenting cleavage specificity for particular MMP isoforms may also be facilitated by other proteomic approaches to further unravel their biological implications. For instance, if an MMP cleavage terminus from a semitryptic peptide can be ascribed to a particular

**Figure.** Peptide-based proteomic workflow. A proteomic analysis typically involves a series of steps beginning with isolation and separation of proteins from a biological source of interest, ultimately leading to resolution of proteome components assigned by bioinformatic mass spectral interpretation. Separation methods, including gel electrophoresis or multidimensional liquid chromatography approaches, are applied to enhance proteome resolution. Isolated proteins are subject to proteolytic digestion, generating peptides amenable to mass spectrometry. Insertion of an additional substrate-guided protein cleavage step has proven useful in facilitating the detection and identification of the innate degradome. This is accomplished by exploiting the natural substrate repertoire, together with standard proteolysis, to overcome potential hindrances imposed by the presence of large peptides and lack of protease sequence specificity.

**Figure.** Peptide-based proteomic workflow. A proteomic analysis typically involves a series of steps beginning with isolation and separation of proteins from a biological source of interest, ultimately leading to resolution of proteome components assigned by bioinformatic mass spectral interpretation. Separation methods, including gel electrophoresis or multidimensional liquid chromatography approaches, are applied to enhance proteome resolution. Isolated proteins are subject to proteolytic digestion, generating peptides amenable to mass spectrometry. Insertion of an additional substrate-guided protein cleavage step has proven useful in facilitating the detection and identification of the innate degradome. This is accomplished by exploiting the natural substrate repertoire, together with standard proteolysis, to overcome potential hindrances imposed by the presence of large peptides and lack of protease sequence specificity.
MMP isoform, this peptide may in turn serve as a selective activity biomarker in interstitial proteomics multiple-reaction monitoring experiments, much as the technique has been applied to the search for cardiovascular disease markers in human plasma. Complementing a semitryptic peptide strategy, top-down proteomic approaches, already in service for cardiovascular biomarker detection, may also prove useful for definitive mapping of full-size MMP cleavage products, including their associated posttranslational modifications. Of note, a number of these high-molecular-weight MMP cleavage peptides may harbor as-yet undiscovered bioactive properties, as suggested by the antiangiogenic effects of arresten, a C-terminal cleavage peptide derived from collagen α-1 (IV), or the antiproliferative and antiangiogenic effects of endostatin, the aforementioned product of MMP-7–mediated collagen α-1 (XVIII) C-terminal cleavage. The expanded set of proteome-resolved cleavage sites reported by Stegemann et al thus represents an initial step in securing a working database for functional interpretation of MMP isoform-specific ECM remodeling.

Identifying human vascular substrates of representative isoforms for 3 major MMP classes was the primary objective of this study. Achieving this goal required development of a strategy to overcome hurdles raised by large ECM peptides lacking terminal amino acid sequence specificity. By modifying a standard proteomic workflow (Figure), Stegemann et al have advanced the capacity to survey interstitial proteolytic landscapes, enhancing resolution of MMP substrate and cleavage site-specific information and thereby expanding this venue of human vascular research. Although contemporary proteomics involves high stringency to ensure accurate peptide and protein assignment, this study further demonstrates the technical versatility required of the field, in that a fit-for-purpose diversity of tools may be interchangeably applied to address unique proteomic properties of the particular biological question at hand. With further expansion and refinement of the resource of MMP-dependent ECM fragments, the ultimate goal will be to use this compendium in targeted experiments to address the consequences of normal and dysregulated MMP interstitial activity. This will necessitate a cyclic return from molecular specificity to systematic physiological function, possibly with iterative incorporation of additional proteomic and other high-throughput data. As demonstrated by other extracellular proteomic and combined metabolomic/proteomic analyses, high-throughput identification and characterization of relationships between components of the extracellular domain can be critical in defining functional transitions between cellular states and cell fate decisions. Systems biology approaches offer an integrative approach by which to achieve data inclusivity and to maximize biological comprehension for disease diagnosis, determination of disease predilection, and assessment of therapeutic modalities, and thus are ideally suited to addressing the biological complexity associated with future investigations into MMP isoform-dependent human ECM remodeling.

Sources of Funding
This study was supported by the National Institutes of Health, Leducq Fondation, Marriott Heart Disease Research Program, and Mayo Clinic. Dr. Terzic is the Marriott Family Professor of Cardiovascular Research at Mayo Clinic.

Disclosures
Dr. Terzic is chair, Council on Functional Genomics and Translational Biology, American Heart Association. Dr. Arrell has no conflicts to report.

References
Substrate-Guided Proteomics Enhances Degradome Resolution
D. Kent Arrell and Andre Terzic

doi: 10.1161/CIRCGENETICS.111.000031
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/6/1/7

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics 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: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org/subscriptions/