Systems Biology – Opportunities and Challenges: 
The Application of Proteomics to Study the 
Cardiovascular Extracellular Matrix

Javier Barallobre-Barreiro¹, Marc Lynch¹, Xiaoke Yin¹, Manuel Mayr¹.
¹King’s British Heart Foundation Centre, King’s College London, London, UK

Short title: Proteomics of the Cardiovascular Extracellular Matrix
Word count: 8775
Table: 1 Figures: 5

Correspondence: Professor Manuel Mayr, King’s British Heart Foundation Centre, 
King’s College London, 125 Coldharbour Lane, London SE5 9NU, UK, Fax: 
+44(0)207848 5298; Tel: +44(0)20 7848 5446; Email: manuel.mayr@kcl.ac.uk

© The Author 2016. Published by Oxford University Press on behalf of the European Society of Cardiology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
ABSTRACT

Systems biology approaches including proteomics are becoming more widely used in cardiovascular research. In this review article, we focus on the application of proteomics to the cardiac extracellular matrix. Extracellular matrix remodelling is a hallmark of many cardiovascular diseases. Proteomic techniques using mass spectrometry provide a platform for the comprehensive analysis of extracellular matrix proteins without a priori assumptions. Proteomics overcomes various constraints inherent to conventional antibody detection. On the other hand, studies that use whole tissue lysates for proteomic analysis mask the identification of the less abundant extracellular matrix constituents. In this review, we first discuss decellularization-based methods that enrich for extracellular matrix proteins in cardiac tissue, and how targeted mass spectrometry allows for accurate protein quantification. The second part of the review will focus on post-translational modifications including hydroxylation and glycosylation and on the release of matrix fragments with biological activity (matrikines), all of which can be interrogated by proteomic techniques.

Keywords:

Cardiovascular diseases, extracellular matrix, post translational modifications, proteomics, systems biology
1. INTRODUCTION

Proteomic techniques using mass spectrometry (MS) provide a platform for the comprehensive analysis of proteins, thereby facilitating the implementation of systems biology approaches and circumventing the limitations of a traditional, reductionist approach adopted by techniques like Western blotting that are based on a priori assumptions of the proteins to be investigated. Furthermore, proteomics is without the constraints of antibody-dependent protein detection and has the capability of detecting post-translational modifications (PTMs), which is beyond the means of gene expression platforms.

Tissue fibrosis is a hallmark of most cases of cardiovascular disease (CVD) and includes modification and deposition of extracellular matrix (ECM). However, detailed studies on cardiovascular ECM have been sparse due to the lack of analytical tools that facilitate comprehensive characterization of its components. In recent years, proteomics has been successfully applied to study the ECM, providing unprecedented insights into its biology and pathological remodeling. In the present review, we describe the utility of ECM proteomics as applied to cardiovascular research and the potential pitfalls. In addition, we highlight the means to overcome common proteomic challenges and present translational applications of proteomic datasets.

2. THE EXTRACELLULAR MATRIX IN CARDIAC DISEASE

The ECM not only confers mechanical stability, but is also a reservoir for bioactive molecules. Remodelling of the ECM, including quantitative but also qualitative changes in composition, is a hallmark of CVD. Numerous studies have demonstrated that structural, but also non-structural ECM proteins play crucial roles during disease progression and normal cardiac physiology.

Table 1 summarizes important findings in clinical studies as well as in animal models of cardiac disease. Additional studies reported ECM proteins as potential biomarkers for...
different cardiac pathologies; these have been intentionally omitted from the Table, the focus of which are ECM and ECM-associated proteins (i.e. extracellular proteases and non-structural proteins that bind to or regulate ECM) from a functional perspective. Most proteins included in the Table were individually studied using antibodies and loss-of-function models in order to assign their relevance to disease. Proteomics can quantify most of these ECM proteins in a single experiment, leading to the identification of previously unreported links between ECM components in disease. For example, in a recent study we demonstrated that genetic deletion of biglycan was accompanied by an unexpected rise of aggrecan in murine aortas.

3. “IN ANTIBODIES WE TRUST”

Until recently, the identification of proteins in a tissue or a protein lysate has been limited by the availability of antibodies that recognise certain portions (epitopes) of a protein of interest. Antibodies have been and continue to be an important component of the armamentarium for protein research but they are not without limitations. While antibody arrays overcome the restriction to only one protein, ECM proteins tend to be under-represented in arrays. The main issues of antibody-based protein quantification, however, remain the same: 1) Usually only a small portion of the protein (epitope) is recognised by an antibody. Protein detection by antibodies relies on the presence of unmodified epitopes. In ECM proteins, however, common PTMs include hydroxylation, glycosylation or fragmentation. Due to epitope masking, ECM proteins may not be detectable by antibodies. 2) Antibodies are often not commercially developed to target proteins in species beyond the commonly used mice and rats such as canine and porcine models for myocardial infarction, rabbit and goat models for studies involving atrial fibrillation and sheep as models of dilated cardiomyopathy. While some anti-human or anti-mouse antibodies will cross-react, many others will not recognize their target protein in different species or will display a high degree of non-specific binding. Vice versa, proteins in the bovine serum supplements of cell cultures can be detected not only in the conditioned media but also in the cell lysates.
Contrary to antibodies, proteomics does not rely on recognition of one specific epitope and can be applied across species. Moreover, the use of MS allows for determination of changes that occur at the protein level (i.e. amino acid modifications) (Figure 1). For example, we demonstrated using MS that the C-terminus of decorin, a small leucine-rich proteoglycan, is often cleaved in the left atrium but not in the ventricle. MS data provided an explanation why the use of different antibodies for the same target protein yielded very different results (Figure 2).

4. EXTRACELLULAR MATRIX REVISITED BY PROTEOMICS

Proteomics is the study of the complete protein component of a living organism, tissue or cell and yields unbiased data without a priori knowledge. The workhorse of modern proteomics is the mass spectrometer and although it is not a new technology per se, it was for a long time confined to areas outside the biological sciences. However, it was the advent of matrix-assisted laser desorption ionization (MALDI) and in particular of electrospray ionization (ESI) - which enables liquid chromatography (LC) systems to be interfaced directly to mass spectrometers - that MS branched from analytical chemistry into biology.

The gold standard for contemporary proteomics is LC-tandem MS (LC-MS/MS). Briefly, the LC column separates the peptides (typically generated by digesting proteins with trypsin) in the analyte prior to ionisation and subsequent MS analysis. In addition to recording the mass of the peptide ions, MS/MS technologies induce the subsequent fragmentation of these precursor ions. The masses of these fragment ions can therefore be used to delineate the amino acid sequence of the peptide. The availability of annotated protein sequence databases and algorithms that match the observed MS/MS spectra to protein entries have been crucial for the biomedical application of MS to study proteins. MS data can also be aligned to databases generated using DNA or RNA sequences to infer amino acid sequences. Current MS technologies now allow for the characterization of the ECM composition and turnover in CVD in unprecedented detail that is not possible using other techniques.
5. EXTRACTION OF EXTRACELLULAR MATRIX PROTEINS

Over the past years our group has focused on ECM remodelling in cardiac\textsuperscript{2,52,60,61} and vascular tissues\textsuperscript{3,4,57,59} using proteomics. In a previous review, we highlighted the potential of proteomics when applied to systems biology\textsuperscript{62}. A recent review by Chang \textit{et al} has focused on clinical applications of ECM proteomics (i.e. biomarker discovery and tissue engineering)\textsuperscript{63}. In this review, we discuss how MS can be used to assess ECM composition in CVD.

Contemporary mass spectrometers have exceptional sensitivity, providing detection at attomole concentrations\textsuperscript{64}. However, such sensitivity is largely confined to pure solutions and not yet achievable in complex biological samples. For example, the dynamic range of proteins present in plasma spans 10 orders of magnitude (e.g. 4x10\textsuperscript{10} pg/ml for albumin compared to few pg/ml for some interleukins)\textsuperscript{65}. Current MS instrumentation can only resolve 4-6 orders of magnitude. While proteomics offers a comprehensive analysis of high abundant proteins it has not yet overcome the difficulties of analysing low abundant proteins in complex samples. Unlike PCR or antibody-based techniques, proteomics lacks the ability to amplify low abundant proteins to aid detection and instead relies on enriching the target proteome.

For instance, cardiac ECM proteins are markedly less abundant than cytosolic and mitochondrial proteins\textsuperscript{46}. Thus, studies that use whole tissue lysates for proteomic analysis inevitably mask the identification of the less abundant ECM constituents. With cardiac tissue this is exacerbated due to the higher cellular content\textsuperscript{2}. Accordingly, methods that enrich for ECM proteins have received considerable interest of late and principally focus on removing plasma contaminants and soluble cellular proteins\textsuperscript{57,66}.

While the inherent insolubility of many ECM proteins lends itself to effective enrichment by decellularization, subsequent proteomic analysis requires all proteins to be solubilized. Standard lysis buffers are not effective for ECM solubilisation. Instead, we implemented a stepwise extraction of vascular ECM proteins\textsuperscript{57}. This involves treating vascular tissues with sodium chloride (NaCl) to remove plasma proteins and extract loosely bound extracellular
proteins before decellularizing the tissue with sodium dodecyl sulphate (SDS). Each incubation step takes 4 hours. Solubilisation of mature ECM proteins is finally achieved by treatment with guanidinium hydrochloride (GuHCl) which destabilizes the ionic, disulfide-dependent protein conformation in large aggregating proteoglycans (versican, aggrecan, etc), small proteoglycans (decorin, biglycan, etc), cell-attachment glycoproteins such as type VI collagen, fibronectins, laminins and basement membrane components. The method was later adapted for the use in porcine cardiac tissue by reducing the incubation time for NaCl and prolonging the SDS treatment (Figure 3A). In smaller animal models (i.e. mouse, rat) cardiac cellularity is proportionally higher compared to that of larger animal models such as pig or goat (Figure 3B). With increased cellularity, decellularization is more difficult to achieve and may require additional enrichment steps, i.e. for glycoproteins or glycopeptides.

Others have adopted similar workflows to extract ECM proteins in a number of tissues. Of note is the Texas 3-step extraction method by Lindsey’s group. In their method, applied to mouse hearts, a similar sequential extraction consisting of NaCl and GuHCl extraction steps as well as the SDS decellularization are performed. In addition, the Texas 3-step method includes further extraction of the insoluble protein pellet after incubation in GuHCl for 48 hours. Notably, the vast majority of ECM proteins are identified in the GuHCl fraction. The pellet, however, contains few polymerized proteins, which are not extracted by our “English Quickstep” method (Figure 3C).

In a recent proteomics study, Johnson TD et al studied the human cardiac ECM. Decellularization was achieved after perfusion with high SDS concentration (i.e. 10 times greater than that used in our protocol) for more than 3 days. This yields a simplified ECM amenable as a biomaterial, but the ECM proteins will be denatured and ECM-associated proteins will be lost during prolonged incubation with such a high concentration of detergents. The study of ECM using MS approaches described below, requires a gentler extraction method that strikes a balance between removal of cellular components while preserving ECM-associated proteins.
6. DISCOVERY PROTEOMICS

Discovery proteomics refers to the use of proteomics as a hypothesis-free tool to globally profile the proteome of a given sample. In discovery proteomics, bottom-up or shotgun proteomics is based on the analysis of peptides generated after enzymatic digestion of a protein mixture. Digested peptides are separated by LC before MS/MS analysis. In comparison to gene expression analysis ECM proteomics offers certain advantages. First, many diseases manifest over years. Therefore, although transcript levels provide a window into cellular activity at the time of harvest, they merely provide an indirect assessment of protein synthesis at a single time point. When studying dynamic entities such as the ECM, transcript levels become more extraneous, particularly as nascent ECM proteins are incorporated into the existing matrix, and actual ECM protein abundance is determined by the balance of protein synthesis, deposition and degradation.

There are multiple MS approaches that can be applied to yield accurate quantitation. However, each approach comes with distinct trade-offs. Label-free methods can provide relative quantification in simpler mixtures. In complex mixtures, isotopic labelling should be employed, which allows multiplexing of samples. For instance, stable isotope labelling with amino acids in cell culture (SILAC) is based on metabolic labelling of proteins in vitro with amino acids containing heavy (e.g. $^{13}$C) stable isotopes. Fully labelled SILAC mice have also been generated. Methods for protein labelling are based on the use of isobaric tags, such as isobaric tags for relative and absolute quantitation (iTraq) or tandem mass tag (TMT). Isobaric tags have the same chemical structure but different isotope substitutions. When samples are labelled with different tags, they can be subsequently mixed in equal portions, and the protein abundance from the different samples can be assessed by comparing the abundance of peptides labelled with the different tags in a single LC-MS/MS run. Although these methods overcome issues such as technical reproducibility, labelling is only introduced after protein extraction and therefore, unlike SILAC, isobaric tags do not allow for in vivo labelling but have been used for quantitative comparisons using tissue samples.
7. TARGETED PROTEOMICS

The discovery proteomics approach is largely limited by the scan speed as peptides are selected for fragmentation based on abundance. This stochastic process results in a bias towards the more abundant proteins. In contrast to discovery proteomics, targeted proteomics focuses on specific proteins. This is useful when a predetermined group of targets are of interest (e.g. ECM proteins). Proteotypic peptides unique to these proteins are quantified in what is known as selected reaction monitoring (SRM) or multiple reaction monitoring (MRM). The targeted approach increases selectivity, sensitivity and accuracy and enables simultaneous measurement of hundreds of transitions in a single LC-MS/MS run. The transitions for proteotypic peptides will be interrogated as a surrogate of total protein levels, but peptides not included in the search (e.g. non-annotated PTMs) are not detected.

This approach is particularly useful to detect CVD biomarkers, as Domanski et al demonstrated in a study that also included ECM biomarkers of fibrosis. Moreover, targeted proteomics constitutes a robust method to validate findings obtained from discovery experiments (Figure 4).

8. POST-TRANSLATIONAL MODIFICATIONS

ECM proteins are often modified by PTMs, most notably hydroxylation and glycosylation.

8.1. Collagen hydroxylation. Collagens are the major fibril-forming proteins in the ECM and they consist of a basic triple-helical conformation. The triple helix increases molecular stability and provides resistance to tensile stress. Although many types of collagen exist, a consistent pattern can be observed for amino acid sequences of all collagens; each chain contains enriched triplet repeats consisting of the sequence X-Pro-Gly. Prolines within these domains become hydroxylated under the action of prolyl-hydroxylases. Hydroxyprolines provide the substrate for the formation of hydrogen bonds between the adjacent collagen alpha chains. Prolyl-4-hydroxylases and prolyl-3-hydroxylase catalyze the hydroxylation of specific proline residues. The former enzyme reacts on proline with the minimum sequence X-Pro-Gly and
the latter appears to require a Pro-4-Hyp-Gly (Hyp is hydroxyproline) sequence\textsuperscript{78,79}. Hydroxylation is a stable, non-reversible PTM that adds +15.99 Da (i.e. an oxygen atom) to proline. X-Pro-Gly domains are rare in ECM proteins other than collagens, and this confers specificity to the acquisition of this PTM.

Similar to prolyl-hydroxylases, lysyl-hydroxylases catalyze the hydroxylation of lysine, which is critical for collagen stability. The specifics of lysine hydroxylation are beyond the scope of this review, and are discussed elsewhere\textsuperscript{80}. Adding hydroxylation as a variable modification, dramatically improves identification and quantification of collagen levels in disease\textsuperscript{3,68}. Ascorbic acid (vitamin C) is a key cofactor for prolyl-4-hydroxylase, and its deficiency causes defects in collagen assembly\textsuperscript{81}. Inhibition of this enzyme has been shown to affect left ventricular remodeling after myocardial infarction in rats\textsuperscript{82}. In this study only proline:hydroxyproline ratios were assessed. MS provides assessment of hydroxylation with concomitant assignment to specific collagen types.

8.2 Glycosylation. Glycosylation is an enzymatic process through which a glycan is covalently attached to a second biomolecule. Glycosylation is a very common form of PTM of ECM proteins. Attached glycans affect ECM protein structure and function by influencing its folding, solubility, aggregation and/or degradation behaviour\textsuperscript{83}. Indeed, aberrant glycoforms are already approved as biomarkers for cancer\textsuperscript{84}. In cardiac tissue, Montpetit \textit{et al} showed that aberrant glycosylation of extracellular domains alters ion channels activity\textsuperscript{85}.

There are two main glycosylation types in mammals: N-glycosylation occurs at the carboxamido nitrogen on asparagine residues (Asn) of secreted/membrane proteins within the consensus sequence Asn-Xaa-Thr/Ser, where Xaa is any amino acid except for proline\textsuperscript{86}. The second main type of glycosylation is O-glycosylation, in which sugar residues attach to serine and threonine residues (Ser, Thr) or, to a much lesser extent to hydroxyproline and hydroxylysine\textsuperscript{87}. The latter two are particularly abundant in collagens and add an additional level of regulation to collagen biosynthesis. If both present, O-glycosylation occurs after N-
glycosylation. Moreover, O-glycosylation is not restricted to secreted proteins and to date, no consensus sequences have been identified for this PTM. ECM proteins may be extensively modified by addition of N- and O-linked large and repetitive glycosaminoglycans (GAGs) and shorter and diverse N- and O-linked oligosaccharides. Aberrant glycosylation can lead to pathological abnormalities and disease. In the last decade, proteomics has emerged as a powerful platform to characterize glycosylation profiles of ECM proteins, including the cardiovascular field. There are two major strategies that can be used to study glycoproteins (Figure 4):

**Indirect MS methods for determining glycosylation profiles.** A common large-scale strategy utilizes a glycopeptide or glycoprotein enrichment step followed by glycan removal. Glycopeptides are usually enriched using lectins, hydrophilic interaction LC, hydrazide or graphite. The method of choice will determine the type of glycopeptides that will ultimately be enriched. After enrichment, PNGase-F is used to enzymatically remove the glycan moiety from asparagine residues, serving two purposes: Firstly, the core peptide can be analysed without interference from sugars during MS/MS and secondly, PNGase-F via a deamidation reaction converts the asparagine to aspartic acid. This conversion is characterized by a 0.984 Da mass shift that can be detected using MS. Moreover, if the reaction is performed in the presence of H$_2^{18}$O, it instead leads to a 2.99 Da mass shift, indicative for the presence of glycosylation at that position. Using this methodology in rat hearts, Parker et al. identified 1556 N-linked glycosites from 972 protein groups. This study provided information on the changes in glycosylation following ischemia and reperfusion. Enzymatic deglycosylation allows for the separate analysis of the core protein and glycan, but the link between the glycans and peptides is lost.

**Direct MS methods for determining glycan structures.** The combined analysis of the glycan motif (glycomics) and the protein identification (proteomics) forms the field of glycoproteomics. For such analysis, proteins in the sample are first digested into peptides, followed by glycopeptide enrichment using zwitterionic hydrophilic interaction LC (ZIC-
HILIC\textsuperscript{93} or alternative approaches\textsuperscript{91,94}. Recently, the combination of higher energy collision dissociation (HCD) and electron transfer dissociation (ETD) have facilitated direct MS analysis of glycopeptides. HCD fragmentation mostly breaks glycosidic bonds, whereas ETD preserves the glycan attachment and fragments the peptide backbone, providing peptide sequence information\textsuperscript{89}. Direct analysis of intact glycopeptides has rarely been applied in the cardiovascular context. Our first study by Yin \textit{et al} characterised the glycopeptides of secretomes from human endothelial cells\textsuperscript{89}. More recently, we have characterized the glycosylation profile of human cardiac ECM proteins\textsuperscript{52}.

8.3. Reversible PTMs on ECM proteins. Glycosylation and hydroxylation are among the most common PTMs in ECM proteins. Importantly, they constitute non-reversible modifications, but reversible PTMs such as phosphorylation and sulfonation also occur. For example, the transmembrane collagen XVII can be phosphorylated and this mechanism regulates shedding of its ectodomain\textsuperscript{95}. Similarly, phosphorylation of osteopontin inhibits vascular calcification\textsuperscript{96}. In a seminal study by Lundby \textit{et al}\textsuperscript{97} proteomics was used to identify phosphosites on fourteen different rat tissues including hearts. Phosphopeptides were enriched using titanium dioxide. Notably, many previously unrecognized phosphosites were reported in ECM proteins including several collagens and non-collagenous ECM proteins such as laminins, fibronectin, versican and decorin to name a few. This methodology was effective on fresh animal tissues, but has yet to be applied to the context of cardiac disease. Challenges to its application will include preservation of short-lived PTMs in patient samples and during sample preparation.

9. FRAGMENTATION OF EXTRACELLULAR MATRIX PROTEINS

Proteolytic fragmentation of ECM proteins by secreted proteases controls their localisation, activation and interaction, adding an additional layer of regulation for tissue processes. Using experimental data, databases/algorithms such as MEROPS and PROSPER have been created to calculate probability matrices for target protease sequences\textsuperscript{98-100}. This is a valuable
resource for research and is particularly useful when used in conjunction with proteomics. Standard proteomics pipelines work with digested protein mixtures (i.e. trypsin digestion) to screen for the abundance of proteins in a tissue. Trypsin cleaves C-terminally to lysine (Lys, K) or arginine (Arg, R) residues. However, proteases other than trypsin can be present in samples and the endogenous proteolytic activity may give rise to non-tryptic peptides. In a study from Stegemann et al, we described a number of proteolytic targets for various MMPs in the vasculature after addition of these proteases to vascular tissue explants.59

When searching for protease targets, appropriate controls are needed (i.e. healthy or non-digested tissues) in order to avoid reporting artefactual cleavages that may arise from experimental processing, or identifying those that are part of normal physiological turnover. Moreover, the addition of broad-spectrum protease inhibitors during extraction reduces the chance of producing artefactual fragmentation. More sophisticated methods include free C- or N-terminal labelling of endogenous protease-generated fragments prior to digestion for MS analysis.101 For example, the TAILS proteomics approach (isotope-based N-terminal labelling) has been successfully applied by Prudova et al. to analyse the degradome of MMP-2 and MMP-9.102 The same authors used a similar methodology to characterize the degradation of proteolytic fragments in human platelets.103 Ultimately, after identification of cleavage sites, targeted proteomics can be used to study the abundance of ECM fragments in clinical samples.

Specific biological activities have been attributed to certain ECM proteolytic fragments (Figure 5).5,19,52,104-122 The term matrikines has been proposed for these fragments. This should not be confused with the term matricryptins, which is more accurately applied to ECM protein domains that are unexposed (and therefore inactive) unless the protein is subject to fragmentation-derived conformational changes. For example, C-terminal cleavage of collagens XV and XVIII, generates restin and endostatin, respectively. Both fragments exert anti-angiogenic activity in vivo.119 Other collagen types also generate biologically active fragments, e.g. collagens IV and VI, which are highly expressed in the cardiac ECM, as
reviewed elsewhere. The large proteoglycan versican is cleaved by proteases from the MMP and ADAMTS families. Versikine is generated by N-terminal cleavage of versican by ADAMTS-1/4, and influences cell proliferation and apoptosis locally. Endorepellin, a C-terminal peptide from perlecan exerts anti-angiogenic effects. Last, non-structural ECM proteins also release fragments, e.g. the small leucine-rich proteoglycan decorin releases decorunt and other fragments, that exert local regulatory roles over cytokines and growth factors. Recently, we demonstrated that decorin is fragmented in the cardiac ECM. We detected C- and N-terminal non-tryptic cleavage sites on decorin by MS. The resulting cleavage products may regulate growth factor availability. Using similar approaches, the Lindsey group identified cleavage products derived from collagen I that promote scar formation after MI.  

10. CONCLUSIONS

The application of MS constitutes one of the biggest technological advances introduced to protein research. It offers an unbiased platform to analyse global protein expression and holds potential in facilitating novel insights. As recently highlighted in a scientific statement of the American Heart Association; it is anticipated that proteomics research will further our understanding of mechanisms of CVD with one important aspect being the elucidation of ECM composition in healthy and diseased cardiovascular tissues. To achieve this goal, bioinformatics approaches should be applied for interpreting the protein datasets and extract the biologically relevant information. Undoubtedly, the amount of data generated by proteomics represent an analytical challenge. In this regard, special attention should be paid to ECM fragments as they hold potential for two purposes: from a diagnostic perspective, they leak from tissues and when released into the blood stream can be used as biomarkers for CVD. Secondly, since many ECM fragments are biologically active, they not only hold enormous potential as therapeutic targets but also as modifiable therapeutic agents – to date an underexplored avenue of cardiovascular medicine.
FUNDING

MM is Senior Fellow of the British Heart Foundation (FS/13/2/29892). The study was supported by an excellence initiative (Competence Centers for Excellent Technologies – COMET) of the Austrian Research Promotion Agency FFG: “Research Center of Excellence in Vascular Ageing – Tyrol, VASCage” (K-Project number 843536) and the NIHR Biomedical Research Center based at Guy’s and St Thomas’ National Health Service Foundation Trust and King’s College London in partnership with King’s College Hospital.
REFERENCES


13. Tao G, Levay AK, Peacock JD, Huk DJ, Both SN, Purcell NH, Pinto JR, Galantowicz ML, Koch M, Lucchesi PA, Birk DE, Lincoln J. Collagen XIV is important for growth and structural integrity of the myocardium. *J Mol Cell Cardiol* 2012;53:626-638.


magnetic resonance imaging in a canine myocardial infarction model. *J Am Coll Cardiol* 2004;44:2383-2389.


82. Nwogu JI, Geenen D, Bean M, Brenner MC, Huang X, Buttrick PM. Inhibition of collagen synthesis with prolyl 4-hydroxylase inhibitor improves left ventricular function and alters the pattern of left ventricular dilatation after myocardial infarction. *Circulation* 2001;104:2216-2221.


Figure 1. Antibody limitations. Detection by antibodies relies on binding to specific regions (epitopes) of the target protein. PTMs such as glycosylation or fragmentation may hinder epitope accessibility. Antibody A recognises permanently non-glycosylated regions and always yields detection independently of sugar removal (left panel). Antibody B recognises epitopes in the vicinity of potentially glycosylated regions. Therefore, recognition is only achieved after deglycosylation (DG). If fragmentation occurs, only antibody C, which recognises an intact region, reveals a degradation pattern. Antibody D targets a region affected by fragmentation and can only detect the intact epitope. Consequently, information about degradation is missed. Proteomics interrogates peptides across the whole sequence and allows for consideration of variable modifications at the amino acid level. Different protein forms can therefore be identified and quantified.

Figure 2. MS to explain discrepancies between different antibodies. An antibody against a C-terminal epitope (green on left panels, red on right panels) detects a markedly lower abundance for decorin in the atrium. An antibody against a different epitope (green on right panel) shows no differences in abundance. This is explained by cleavage of decorin at the C-terminus, which was detected in the atrium using MS.

Figure 3. Enrichment of cardiac ECM proteins. A) Our previously published 3-step ECM extraction method for cardiac tissue is based on decellularization and ensures enrichment and detection of ECM proteins. The image shows a decellularized heart after prolonged SDS perfusion. The ECM is solubilized by guanidine hydrochloride (GuHCl) and analysed using proteomics. B) Smaller species display higher levels of cardiac cellularity as measured by the ratio of 3 members of different ECM protein classes and the cardiac-specific troponin T. C) Proteins identified in murine hearts using the Texas 3-step extraction method compared to
those identified by our previously published method (see Drozdov et al\textsuperscript{61}). Most proteins are identified by both methods. Unlike the Texas 3-step method, our “English Quickstep” method did not include an analysis of the remaining pellet after GuHCl extraction.

**Figure 4. MS strategies for ECM characterisation.** Untargeted proteomics is appropriate for discovery experiments where no \textit{a priori} information is available. When a delimited number of targets of interest are known \textit{a priori}, targeted proteomics offers a robust method for detection and quantification. Novel MS methods such as a combination of HCD and ETD allow for characterisation of complex PTMs including glycosylation.

**Figure 5. Biological activity of ECM fragments.** Fragments derived from a variety of ECM proteins (i.e. matrikines) exert functions that regulate diverse cellular and tissue processes. Proteomics offers a tool for the analysis of known ECM fragments as well as the discovery of previously unknown fragments with functions potentially important for cardiac physiology and putative therapeutic targets. * Indicates putative fragments with activities only characterised after exogenous administration.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Clinical context</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adiponectin</td>
<td>Cardiac remodelling (m)</td>
<td>Induces cell migration, MMP activation, and collagen remodeling via APPL1-AMPK signalling</td>
</tr>
<tr>
<td>ADAMTS9</td>
<td>Developmental defects (m)</td>
<td>Haploinsufficiency leads to reduced versican cleavage, associated with cardiac anomalies.</td>
</tr>
<tr>
<td>Biglycan</td>
<td>MI (m)</td>
<td>Required for adaptive remodelling</td>
</tr>
<tr>
<td>Cathepsin-K</td>
<td>AF (h, rb)</td>
<td>Increased levels and activity accompanied atrial changes linked to the AngII/ATR1 signaling pathway</td>
</tr>
<tr>
<td>Cathepsin-S</td>
<td>MI (m)</td>
<td>Mediates fibroblast trans-differentiation during remodelling</td>
</tr>
<tr>
<td>Collagen I</td>
<td>Dilated cardiomyopathy (m)</td>
<td>Point mutation induces cardiomyopathy</td>
</tr>
<tr>
<td>Collagen VI</td>
<td>MI (m)</td>
<td>Absence improves cardiac function, structure, and remodelling</td>
</tr>
<tr>
<td>Collagen XIV</td>
<td>Developmental defects (m)</td>
<td>Important for growth and structural integrity of the myocardium</td>
</tr>
<tr>
<td>Collagen XV</td>
<td>Hypertension (m)</td>
<td>Necessary for remodelling. Deficiency predisposes to cardiomyopathy</td>
</tr>
<tr>
<td>Connective tissue growth factor</td>
<td>Pressure overload (m)</td>
<td>Inhibition attenuates left ventricular remodeling and dysfunction</td>
</tr>
<tr>
<td>Decorin</td>
<td>Left ventricular assist device implantation (h)</td>
<td>Ameliorates adverse remodeling by mediating transforming growth factor-beta inhibition</td>
</tr>
<tr>
<td></td>
<td>MI (m)</td>
<td>Absence leads to abnormal scar tissue formation</td>
</tr>
<tr>
<td>Fibronectin</td>
<td>MI (m, h)</td>
<td>Essential for progenitor cell response during cardiac repair</td>
</tr>
<tr>
<td></td>
<td>MI (m)</td>
<td>Lack of EDA domain promotes survival and prevents adverse remodelling</td>
</tr>
<tr>
<td>Fibulin-2</td>
<td>MI (m)</td>
<td>Loss protects against progressive ventricular dysfunction</td>
</tr>
<tr>
<td>Laminin alpha-4</td>
<td>Dilated cardiomyopathy (h, z)</td>
<td>Mutations cause human cardiomyopathy via defects in cardiomyocytes and endothelial cells</td>
</tr>
<tr>
<td>Lumican</td>
<td>Hypertrophy (m)</td>
<td>Deficiency results in cardiomyocyte hypertrophy with altered collagen assembly</td>
</tr>
<tr>
<td>Protein</td>
<td>Condition/Model</td>
<td>Effect/Function</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mimecan</td>
<td>MI (m, h)</td>
<td>Prevents cardiac dilatation and dysfunction collagen strengthening(^{23})</td>
</tr>
<tr>
<td>MMP-14</td>
<td>Pressure overload (m)</td>
<td>Mediates pro-fibrotic signalling, leading to alterations in interstitial fibrosis and diastolic function(^{24})</td>
</tr>
<tr>
<td>MMP-28</td>
<td>MI (m)</td>
<td>Deletion exacerbates cardiac dysfunction and rupture by inhibiting M2 macrophage activation(^{25})</td>
</tr>
<tr>
<td>TIMP-2</td>
<td>Pressure overload (m)</td>
<td>Loss leads to exacerbated left ventricular dysfunction and adverse ECM remodeling(^{26})</td>
</tr>
<tr>
<td>MMP-9</td>
<td>AF (p, h)</td>
<td>Increased gelatinase activity contributes to atrial ECM remodelling(^{27,28})</td>
</tr>
<tr>
<td>MMP-2</td>
<td>MI (h,m)</td>
<td>Crucial for generation of bioactive collagen I fragments that promote scar formation after MI(^{15})</td>
</tr>
<tr>
<td></td>
<td>MI (m)</td>
<td>Deletion leads to decreased collagen accumulation and left ventricular enlargement(^{29})</td>
</tr>
<tr>
<td>Osteopontin</td>
<td>MI (m, r)</td>
<td>Contributes to ischemia-reperfusion injury, and deletion/inhibition prevents cardiac rupture(^{30,31})</td>
</tr>
<tr>
<td>Periostin</td>
<td>MI (r)</td>
<td>Blockade of Exon 17 preserves cardiac performance(^{33})</td>
</tr>
<tr>
<td></td>
<td>Diabetic cardiomyopathy (r)</td>
<td>Elevated expression of periostin in diabetic cardiomyopathy and the effect of valsartan(^{34})</td>
</tr>
<tr>
<td>Perlecan</td>
<td>Developmental defects (m)</td>
<td>Perlecan is critical for heart stability(^{35})</td>
</tr>
<tr>
<td>SPARC</td>
<td>MI (m)</td>
<td>Mediates early ECM remodeling(^{36})</td>
</tr>
<tr>
<td>Tenascin-C</td>
<td>Pressure overload (m)</td>
<td>Accelerates fibrosis by activating macrophages via the integrin αVβ3/nuclear factor-κB/interleukin-6 axis(^{37})</td>
</tr>
<tr>
<td></td>
<td>MI (m)</td>
<td>May aggravate left ventricular remodelling and function(^{38})</td>
</tr>
<tr>
<td>Thrombospondin-1</td>
<td>Pressure overload (m)</td>
<td>Protects myocardium by modulating fibroblast phenotype and matrix metabolism(^{39})</td>
</tr>
<tr>
<td></td>
<td>MI (d, m)</td>
<td>Role in preventing expansion of healing myocardial infarcts(^{40})</td>
</tr>
<tr>
<td>Thrombospondin-4</td>
<td>Pressure overload (m)</td>
<td>Regulates myocardial fibrosis and remodelling(^{41})</td>
</tr>
<tr>
<td>Versican</td>
<td>Developmental defects (m)</td>
<td>Associated with chamber specification, septation, and valvulogenesis in the developing heart(^{42})</td>
</tr>
</tbody>
</table>

**Abbreviations used:** MI, myocardial infarction; AF, atrial fibrillation; m, mouse; h, human; rb, rabbit; z, zebrafish; r, rat; p, pig; d, dog
Glycosaminoglycan (GAG) ▼ N-linked oligosaccharide ▣ Antibody recognition site

Figure 1

Ab Epitope recognition

MS Amino acid sequence

Diffuse bands due to glycans
Epitope masking due to glycosylation
Epitope masking due to cleavage

Proteins detected regardless of size
Glycosylated & non-glycosylated
Tryptic and non-tryptic peptides
Figure 2

C-terminal DCN Ab (green)
Internal DCN Ab (green)
C-terminal DCN Ab (red)

Left ventricle
Full-length DCN

Full-length and truncated DCN

Predominant ventricular form
Full-length decorin

Predominant atrial form
C-terminal truncation detected by MS

F330 ↑ S331
A. Cardiac tissue

Decellularization

1h NaCl → 16h SDS → 72h GuHCl

ECM analysis

Heart 1

Decorin
Biglycan
Mimican

Heart 2

kDa
-80
-45
-34

B. Proteomics of the ECM from murine hearts

Texas 3-step
(Ref. 66)

“English Quickstep”
(Ref. 61)

C. Proteomics of the ECM from murine hearts

CO6A3
EMIL1
POSTN

N=63

CO2A1
CO6A4
CO7A1
FBN2
CO15A1
LAMA1
CO24A1
LAMA3
CO27A1
LAMA4
CO11A2
MFAP5
Extracts enriched in ECM proteins

Deglycosylation
PNGase F + H$_2^{18}$O

Glycopeptide enrichment
ZiC-HILIC

Untargeted

Targeted

HCD pd ETD
HCD alt ETD

Peptide identification/ quantification
Indirect glycopeptide identification
Peptide quantification
Direct glycopeptide characterisation

Top-N MS/MS

Unmodified peptides (i.e. non-glycosylated)

Modified peptides (i.e. PNGase-F treatment)

HCD - ETD

High mannose
Complex
Hybrid

Targeted

MRM / SRM

Modified peptides (i.e. PNGase-F treatment)

(Deam)+3
Figure 5

ECM bioactive fragments

- CO1A1 (C-1158/59)
- COA1 (restin)
- CO4A1 (arresten)
- PGBM (endorrepelin)
- CO1A1 (endostatin)
- POST (FAS1*)
- PGCA/HPLN1 (Metastatin)
- CO6A3 (endotrophin)
- ELN (VGVAPG)
- FINC (EDA-domain*)
- TENA (FN3 domain*)
- PGS2 (C-term)
- CO4A1 (arresten)
- HGF (N-term*)
- PGS2 (N-term)
- CO1A1 (C-1158/59)
- FINC (N, C-term)
- CO4A1 (arresten)
- CSPG2 (versikine)
- TENA (EGF domain)
- VTNC (N-term)
- OSPT (MMP-cleaved)
- LAM332 (MMP2-cleaved)

Enhances

Inhibits

Angiogenesis
- Hypertrophy
- Dilatation
- Proliferation
- Apoptosis
- Adhesion
- Migration