Functional Insights from the Proteomic Inventory of Ovine Forestomach Matrix

Sandi G. Dempsey, † Christopher H. Miller, † Ryan C. Hill, † Kirk C. Hansen, † and Barnaby C. H. May ††

† Aroa Biosurgery Limited, Airport Oaks, Auckland 2022, New Zealand
‡ Omix Technologies LLC, Bioscience 1, 12635 E. Montview Blvd. Suite 100, Aurora, Colorado 80045, United States

Supporting Information

ABSTRACT: Ovine forestomach matrix (OFM) is a decellularized extracellular matrix (dECM) biomaterial that serves as a scaffold for remodeling damaged soft tissue. dECMs are used in a variety of clinical applications, and their regenerative capacity is encoded not only in their biophysical properties but also in their molecular diversity. In this study, the proteome of OFM was characterized via both targeted and global mass spectrometry (MS) with the use of heavy isotope labeled (SIL) internal standards. Proteins were identified following either chemical digestion or extraction using saline or guanidine hydrochloride, followed by high resolution size exclusion chromatography. Identified proteins were annotated using the matrisome database and molecular function using the gene ontology database. The characterization identified 153 unique matrisome proteins, including 25 collagens, 58 glycoproteins, 12 proteoglycans, 13 ECM affiliated proteins, 20 ECM regulators, and 23 secreted factors. This inventory represents a comprehensive array of matrix proteins that are retained in OFM after processing. The diversity of proteins identified may contribute to OFM’s remodeling capacity in clinical applications.

KEYWORDS: proteome, matrisome, extracellular matrix, ovine forestomach matrix, QconCat, mass spectrometry

INTRODUCTION

In living tissue, cells cooperate with their surrounding extracellular matrix (ECM) and vice versa via “dynamic reciprocity”: a process that is critical for tissue development, homeostasis, and tissue repair. In every stage of soft tissue regeneration a dynamic relationship exists between cells and the various structural, adhesion, signaling, and metabolic proteins of the ECM. ECM proteins from different types of tissue have been catalogued in a “matrisome” database. The most abundant proteins in the ECM are structural collagens, which form a fibrous tissue network and confer strength and structure to living tissue. However, there is a vast number of less abundant proteinaceous ECM components, all with different roles in tissue homeostasis and the remodeling of living tissue.

Suitably designed biomaterials for soft tissue regeneration have found clinical applications where normal tissue ECM is missing or damaged by disease. To serve as biomimetics of normal tissue ECM, these biomaterials need to contain both structural and biological queues in order to effectively participate in the dynamic interplay between patient cells and the ECM during tissue regeneration. There is a spectrum of complexity within these biomaterials, ranging from synthetic polymers to natural matrices containing a large number of ECM proteins. Biomaterial scaffolds can be designed using “bottom-up” or “top-down” approaches. In the bottom-up approach, the scaffold is engineered from known components of well-established composition and purity to yield products with defined porosity, strength, and fiber architecture. The “top-down” approach utilizes subtractive manufacturing to remove unwanted components from a source tissue. This approach has been used to produce “decellularized ECMs” (dECMs) from a range of fresh allogeneic and xenogeneic tissues that are processed to remove cellular components and therefore prevent a foreign body response in the recipient. The resultant dECM biomaterials are predominantly comprised of collagen fibrils; however, many other proteins may also be retained and provide a biomimetic substrate to support tissue remodeling and repair. It is thought that the efficacy of dECM-based biomaterials in tissue repair is a result of their biochemical heterogeneity, which mimics that of native tissue ECM.

Ovine forestomach matrix (OFM) is a dECM biomaterial, composed of structural and functional proteins, including collagens I, III, and IV, elastin and glycosaminoglycans (GAGs, e.g., heparan, chondroitin, and hyaluronic acid) as well as minor components, such as fibronectin, fibroblast...
growth factor (FGF2), and laminin. Structural studies have demonstrated that the decellularized biomaterial retains the native structure of tissue ECM. Previous studies have shown that the biology of OFM promotes regeneration of damaged tissue, vasculogenesis, and undergoes constructive remodeling. To date, OFM has been clinically used for the management of acute and chronic wounds and also in abdominal wall repair. It is thought that the clinical success of OFM is based on its composition of ECM proteins and its preserved native protein structure; both of these factors present a scaffold for rebuilding tissue that is not rejected by the host and stimulates cellular processes required for constructive regeneration of tissue.

Modern MS methods enable a more detailed characterization of complex biological mixtures and are being increasingly employed to deconvolute the composition of tissue ECM and the matrisome. Similarly, MS has been employed to catalogue the protein content of a number of dECMs produced from pancreas, lung, bladder, and liver. For example, a characterization of decellularized urinary bladder matrix generated a proteomic inventory including 129 proteins, which was further analyzed to reveal 80 matrisome proteins. In all cases, these analyses have demonstrated that tissue decellularization eliminates a large proportion of nonmatrisome proteins, and that the resultant dECM biomaterials were a rich milieu of structural and functional proteins.

The objective of this research was to characterize the molecular diversity of proteins in OFM to identify important cell–ECM interactions that may occur during constructive remodeling. Given the molecular complexity of OFM, several approaches were undertaken in order to maximize the completeness of the resultant inventory. This was achieved through sequential fractionation of OFM proteins and combining a quantitative LC–MS method targeted at ECM proteins, in addition to a global MS/MS approach. Identified proteins were annotated and compared to the known human matrisome and mapped to tissue repair processes.

## MATERIALS AND METHODS

### General Methods

OFM was prepared from ovine forestomach tissue using proprietary methods (Aroa Biosurgery, Auckland, New Zealand). Completeness of decellularization was determined by measurement of DNA content as previously described. All samples were terminally sterilized by ethylene oxide (EO) prior to sample extraction and analysis.

### QconCAT Design

Stable isotope-labeled QconCATs were designed as previously described. QconCATs were used to make 418 peptides covering 239 proteins in the *Ovis aries* proteome.

### Sample Preparation and MS Analysis

Samples suitable for proteomic analysis were prepared using (1) chemical digestion by cyanogen bromide, (2) saline extraction and size exclusion chromatography (SEC), or (3) guanidine hydrochloride (GdnHCl) extraction and SEC. Analysis of the samples was conducted using liquid chromatography (LC)–MS/MS global proteomic analysis and LC–selected reaction monitoring (SRM) targeted analysis. This gave six data sets, based on experimental schema presented in Figure 1. For SRM analysis of insoluble proteins (Method B), a library of 267 targets was used. For targeted identification of growth factors (Methods E and F) a library of 22 peptides based on ECM associated growth factors was used.

### Cyanogen Bromide (CNBr) Digestion

Samples of OFM were cut to ~1 cm², weighed, and digested with freshly prepared cyanogen bromide buffer (100 mM CNBr/86% TFA (v/v) solution) at 10 mg/mL with agitation at room temperature for 17 h, before three washes with 100 mM ammonium bicarbonate (pH 8.0).

### Saline Extraction

Samples of OFM were weighed and cut into ~1 cm² squares before extraction using a saline extraction buffer of 1 M NaCl, 10 mM Na₂PO₄, 1X Halt Protease Inhibitor (Sigma-Aldrich). Samples were treated in a bath sonicator for 15 min at room temperature and then placed on a stir plate for 24 h at 4 °C. The supernatant was collected and spun at 18 000 g to remove residual insoluble material. A total of 150 mL of the resulting solution was then concentrated using a 3 kDa Amicon filter (Sigma-Aldrich) to ~50 μL. A volume of 350 μL of acetonitrile solution (25% v/v in H₂O) was then added to the filter and vortexed to bring concentrated proteins into solution. No visible particulate remained at this point.
Guanidine Hydrochloride (GdnHCl) Extraction

Samples of OFM were weighed to approximately 4 g and cut into ~1 cm² squares before extraction with GdnHCl extraction buffer (4 M GdnHCl, 50 mM EDTA, 50 mM Tris pH 7.4, and 1× Halt Protease Inhibitor). Samples were sonicated for 15 min in a bath sonicator at room temperature and then stirred on a stir plate for 24 h at 4 °C. The supernatant was collected and spun at 18 000 × g for 30 min to remove residual insoluble material. A total of 150 mL of the resulting solution was then concentrated using a 3 kDa Amicon filter (Sigma-Aldrich) to ~50 μL. A volume of 350 μL of acetonitrile solution (25% v/v in H₂O) was then added to the filter to vortex and bring concentrated proteins into solution. No visible particulate remained at this point.

Size Exclusion Chromatography (SEC)

SEC was carried out on a Zenix-C SEC-100 (3 μm, 100 Å 7.8 × 300 mm) column using a Dionex Ultimate 3000 UHPLC system. Run conditions consisted of running an isocratic elution for 2 column volumes of 25% acetonitrile (ACN), and 0.05% formic acid (FA). A total of 12 fractions were collected and pooled. Sizing fractions were concentrated over a 10 kDa filter. A sample of recombinant human connective tissue growth factor (CTGF) (Sigma SRP4702) was used as a control to optimize SEC conditions and direct the pooling of fractions as it has a similar biochemical properties, such as PI and MW, to many ECM-associated growth factors of interest.

QconCat: Liquid Chromatography–Selected Reaction Monitoring (LC–SRM)

The ¹³C₆ labeled QconCAT standards were added to each fraction at a mass ratio of 1:400. Tryptic digests were carried out via Filter Aided Sample Preparation (FASP). Briefly, samples QconCAT's were added to the filters prior to digestion when applicable and samples were reduced, alkylated and digested with 5 ng/μL sequencing grade trypsin (Promega) overnight at 37 °C. Tryptic digests were acidified with 0.1% FA and peptides were extracted with three subsequent washes. Peptides were concentrated on a speed-vacuum and then brought up to final volume representing 20% of the starting material.

All samples were analyzed by both targeted LC–SRM and global LC–MS/MS. A targeted, scheduled SRM approach was performed using a QTRAP 5500. Each sample (8 μL) was injected and directly loaded onto an Agilent C18 column (Zorbax SB-C18, 5 μm 150 × 0.5 mm) with 5% ACN, 0.1% FA at 30 μL/min for 3 min. A gradient of 5–28% ACN was run for 38 min to differentially elute QconCAT peptides or growth factor targets. The MS was run in positive ion mode with the following settings: source temperature of 200 °C, spray voltage of 5300 V, curtain gas of 20 psi, and a source gas of 35 psi (nitrogen gas). Method building and acquisition were performed using the instrument supplied Analyst Software (Version 1.5.2).

Global Proteomic Analysis: LC–MS/MS

Samples were analyzed on a LTQ Orbitrap Velos MS (Thermo Fisher Scientific) coupled with an Eksigent nanoLC-2D system. On the LTQ Orbitrap Velos/Eksigent system, 8 μL of sample was loaded onto a trapping column (ZORBAX 300SB-C₁₈, 5 μm 3.0 mm, 5 μm) and washed with 2% ACN, 0.1% FA at a flow rate of 10 μL/min for 10 min. The trapping column was then switched online with the nanopump at a flow rate of 600 nL/min. Peptides were separated on an in-house-made 100 μm × 150 mm fused silica capillary packed with Synergi Hydro-RP C₁₈ Resin (Phenomenex; Torrance, CA) over an 85 min gradient from 6–40% ACN. The flow rate was adjusted to 350 nL/min after 10 min to increase the effective separation of the peptides. MS data acquisition was performed using Xcalibur (version 2.1) software. Collision-induced dissociation was used to produce the fragment ions in the linear ion trap from the precursor ions, which were measured in the Orbitrap mass analyzer. For every MS scan, the 20 most intense ions were selected for fragmentation, and masses selected for fragmentation were then excluded for a duration of 120 s after a repeat count of 3.

Bioinformatics

For global proteomics data, RAW files were converted into peak lists using an in-house script (PAVA). Peak lists were then searched against Byonic-Preview to identify any unexpected post-translational modifications (PTM) or search conditions to carry through the remainder of the analysis. Peak lists were then searched using a MASCOT server utilizing the following parameters; (1) Full trypsin and Semitryptic specificity were searched separately, (2) one missed cleavage was allowed, (3) carbamidomethylation on cysteine was defined as a fixed modification, (4) methionine mono/di/trioxygenation, proline hydroxylation, N-terminal glutamate to pyro-glutamine and methionine to homoserine-lactone (Method A only) were defined as variable modifications for the database searches. Result files were loaded into Scaffold (Proteome Software). Results were directly exported from Scaffold into Excel (Microsoft). LC–SRM data was directly loaded into Skyline, and peaks were manually validated. Results included protein name, nmol/g, and coefficient of variance between three replicates.

Annotation and Categorization of proteins

The matrisome annotator tool was used (matrisome.org, accessed September 2018) to categorize proteins into divisions: “core matrisome” or “matrix associated” or “other”, as well as category “collagens”, “ECM glycoproteins”, “proteoglycans”, “ECM-affiliated”, “ECM regulators”, and “secreted factors”.

Proteins that were matched with the matrisome proteins were uploaded to the gene ontology Web site to generate a list of known functional properties for each protein. Functional gene ontology (GO) categories identified in the target list were simplified into four categories listed below:

1. Structure. Any functional gene ontology record indicating “structural components of ECM”. Any notations relating to “strength, elasticity or compression resistance” were also recorded.

2. Binding or Interaction Related Function. Records pertaining to protein, ion, protease, heparin, GAG, growth factor, integrin binding, or cell adhesion were recorded in this category. Notations relating to cell adhesion or integrin binding were highlighted as well as “growth factor binding”.

3. Signaling Function. Proteins with functional GO categories listed as receptor binding, growth factor activity, cytokine, or hormone were listed in this category. Annotations relating to growth factor and cytokine activity were highlighted.

4. Metabolism. Functions with a protease, phosphatase, reductase, and oxidase were listed in this category as well as inhibitors of protease and peptidase activity.
Table 1. Summary of Methods and Outcomes

<table>
<thead>
<tr>
<th>method</th>
<th>MS method</th>
<th>target</th>
<th>quantitative</th>
<th>outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Global LC–MS/MS</td>
<td>Insoluble proteins</td>
<td>No</td>
<td>Identified 26 unique matrisome proteins and 9 nonmatrisome proteins</td>
</tr>
<tr>
<td>B</td>
<td>QconCat LC–SRM</td>
<td>Insoluble proteins</td>
<td>Yes</td>
<td>Identified 36 unique matrisome proteins and 6 nonmatrisome proteins, from a library of 267 proteins</td>
</tr>
<tr>
<td>C</td>
<td>Global LC–MS/MS</td>
<td>Saline-soluble proteins</td>
<td>No</td>
<td>Identified 646 proteins including 121 matrisome proteins and 531 nonmatrisome proteins.</td>
</tr>
<tr>
<td>D</td>
<td>LC–SRM</td>
<td>Saline-soluble proteins</td>
<td>No</td>
<td>Identified 11 matches from 22 targets</td>
</tr>
<tr>
<td>E</td>
<td>Global LC–MS/MS</td>
<td>Chaotrope-soluble proteins</td>
<td>No</td>
<td>Identified 798 proteins including 131 matrisome proteins and 678 nonmatrisome proteins.</td>
</tr>
<tr>
<td>F</td>
<td>LC–SRM</td>
<td>Chaotrope-soluble proteins</td>
<td>No</td>
<td>Identified 3 matches from 22 targets</td>
</tr>
</tbody>
</table>

Table 2. Summary of Matrisome Proteins Identified in OFM by Category

<table>
<thead>
<tr>
<th>method</th>
<th>collagens</th>
<th>glycoproteins</th>
<th>proteoglycans</th>
<th>ECM-affiliated</th>
<th>ECM regulators</th>
<th>secreted factors</th>
<th>matrisome proteins</th>
<th>other proteins</th>
<th>total proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>26</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>B</td>
<td>13</td>
<td>17</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>C</td>
<td>21</td>
<td>50</td>
<td>11</td>
<td>9</td>
<td>16</td>
<td>14</td>
<td>121</td>
<td>531</td>
<td>646</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>52</td>
<td>11</td>
<td>12</td>
<td>19</td>
<td>15</td>
<td>131</td>
<td>672</td>
<td>798</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

The final inventory of discovered proteins was cross referenced against ECM proteins listed in recent reviews concerning matrikines and cryptic peptides.39–41

RESULTS

OFM samples were processed and analyzed using several methods that included a targeted QconCat method aimed and quantifying ECM proteins, using SIL peptides generated using the QconCAT approach, and targeted LC–SRM method aimed at semiquantitative information on growth factors, and a broader global LC–MS/MS approach (Figure 1). A summary of results for each of the six methods is provided in Table 1. The full data sets from each of the six methods employed is included as Supporting Information.

The global LC–MS/MS analysis (Method A) resulted in the identification of 36 ovine proteins from 7928 spectra. Acceptance thresholds were set to 99% confidence for peptide identifications and 95% for protein identifications with a minimum of 2 peptides per protein. This criteria resulted in a false discovery rate (FDR) for peptide identification at 0.002% and an FDR for protein identifications at 1.2%. The gene list provided in the data set was annotated using the matrisome annotator tool to categorize the proteins. Categorization of the proteins identified using Method A is included in Table 2 and compares the number of protein hits versus the known proteins annotated for the human in silico matrisome.

The targeted QconCat LC–SRM method (Method B) using 657 stable isotope labeled peptides confidently identified 70 peptides belonging to 42 distinct proteins. This data set provided the relative quantity of proteins in nmol/g from three samples, along with the coefficient of variance between replicates (Supporting Information). The gene list provided in the data set was annotated using the matrisome annotator tool (Table 2). Thirty-six proteins were matched from the human matrisome using Method B, while a further six proteins where classified as “other” or nonmatrisome.

In comparison to proteins identified using Method A, Method B identified a further 15 proteins (Figure 2), including collagens (VI, XV, XVIII, and XXI), proteoglycans (lumican and mimecan) and glycoproteins (dermapontin, tenasinC, SPARC, emilin1, fibrillin, nidogen, and osteopontin).

PTM and nontryptic cleavage propensity identified in LC–MS/MS are displayed in Table 3.

Isolation of the proteins using either saline (Method C) or GdnHCl (Method E) extraction greatly improved the variety of the proteins identified (Table 2). For example, isolation of the saline soluble components and analysis via global LC–MS/MS (Method C) identified a total of 652 unique proteins, comprising 122 matrisome proteins and 530 nonmatrisome proteins (Table 2). The saline or GdnHCl extracts were further analyzed by targeted QconCat SRM analysis using a library of 22 proteins of interest, which included growth factors and tissue inhibitors of metalloproteinases (TIMPs) (Methods D and F). This approach identified 11 secreted proteins.
matrisome proteins, including the growth factors BFGF, FGF7/KGF, BMP4, BMP7, BMP9, CSF3, CTGF, EGF, and HGF. As shown in Figure 3, proteins identified using Methods A and B were predominantly from the “collagen” category (89% and 94% respectively) as we expected for the insoluble material. Using SEC (Method C and E) shifted the diversity of identified proteins to include a greater number of other matrisome proteins such as glycoproteins, proteoglycans and ECM associated proteins.

Data generated from Methods A, B, C, and E were used to estimate the abundance of proteins in each sample. For Method A, the relative abundance of each protein is estimated using the number of peptide spectral matches (PSMs) for each unique protein. For Method B the amount of protein is expressed in nmol/g of material. For Methods C and E, the signal intensity and PSM values are calculated by combining the data from all fractions after SEC. Figure 3 illustrates the proportion of each type of core matrix or matrix associated protein identified using separate approaches.

The complete inventory of identified proteins across all six methods was compiled and annotated according to matrisome categories, summarized as Figure 5. Unique proteins identified using all methods were sorted by matrisome category and subcategory/type/family as shown in Figure 5; proteins were color coded based on whether their molecular functions (as described by gene ontology) could be described as “structural” (gray square), “binding interaction” (blue diamond), “signaling” (green circle), or “metabolism” (red star). Within these four categories, functions relating to characteristics of interest for soft tissue remodeling were highlighted by letters within each symbol including strength, elasticity, compression resistance, growth factor activity and cell adhesion and protease inhibition. Proteins with known matrikines or cryptic active sites were annotated, based on current literature on known matrikines derived from ECM proteins,41 these are indicated in Figure 5 with an asterisk (*).

**DISCUSSION**

Significant proteomic diversity is apparent between matrices sourced from different tissue types, healthy versus diseased tissue, and young versus old tissue.30 Top-down dECM biomaterials generated from natural sources have been used in a number of medical applications, and the extent to which proteins from native ECM is retained during the process of decellularization has been widely discussed.5,6,22 Biomaterials derived from dECM have been subject to biochemical characterization; however, given their heterogeneous nature it is not yet possible to comprehensively define their molecular composition, and doing so will contribute to the evolution of such devices and expansion of their clinical potential.

In this study, we aimed to inventory the protein content of OFM, a dECM generated from ovine forestomach tissue, using a proteomic approach. OFM is used clinically in the management of wounds and in soft tissue repair and is terminally sterilized using EO. While EO may introduce covalent modifications to the OFM proteins, EO-sterilized samples were utilized in order to give the most clinically relevant insight into the proteome of OFM. Initially, an ECM-targeted quantitative approach (Method A) was employed in order to generate a snapshot matrisome composition of OFM. Using reference samples of 13C-labeled peptides homologous to peptides within the *Ovine aries* proteome, a list of proteins

<table>
<thead>
<tr>
<th>Cleavage Analysis</th>
<th>prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ragged N-terminal peptide (semitryptic)</td>
<td>20.3</td>
</tr>
<tr>
<td>Ragged C-terminal peptide (semitryptic)</td>
<td>1.8</td>
</tr>
<tr>
<td>Nontryptic peptides</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modifications</th>
<th>prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydroxyproline</td>
<td>74.4</td>
</tr>
<tr>
<td>Methionine Sulfone</td>
<td>5.4</td>
</tr>
<tr>
<td>Methylation</td>
<td>3.4</td>
</tr>
<tr>
<td>All other modifications</td>
<td>&lt;1.8</td>
</tr>
</tbody>
</table>

Table 3. Prevalence of PTMs Identified by Byonic-Preview

![Figure 3. Abundance of proteins identified using different approaches (Method A, B, C, and E) according to matrisome category: “collagens”, “glycoproteins”, “proteoglycans”, “regulators”, “affiliated proteins”, “secreted factors”, and “other”.](image-url)
was generated and annotated giving a much greater understanding of the protein content of OFM than previous targeted approaches described in Lun et al. Using this approach, it was possible to identify 36 proteins in different matrisome categories including 13 collagens, 2 proteoglycans, and 8 glycoproteins. In parallel a global proteomic analysis of this chemical digested sample revealed certain key proteins that homologous QconCAT peptides were not available for, including collagen III, elastin, and laminin. As shown in Figure 3, the targeted approach (Method A) demonstrated

Figure 4. Relative abundance of collagens (left), proteoglycans (center), and glycoproteins (right) using different methods: (A) chemical digestion and global MS (Method B); (B) chemical digestion and SRM (Method A); (C) saline or gnd extraction and SEC (Methods C and E).
that collagen content could be quantified as 89% of the material, which correlates well to biochemical quantification of collagen carried out by Lun et al. A list of expected PTMs identified by LC–MS are shown in Table 3. The most prevalent PTM is hydroxypoline (74.4%). Hydroxyproline is one of the most abundant amino acids in collagen, accounting for ~11% of total amino acid signal in fibrillar collagens alone.83

Figure 5. OFM proteome listed by matrisome category and assigned icons based on functional classified as “structural component” (gray) or “binding interaction” (blue), “signaling” (green) and metabolic/enzymatic or inhibitor function (red). Proteins which are known to contain matrikines or matricryptins are indicated*. DOI: 10.1021/acs.jproteome.8b00908 J. Proteome Res. XXXX, XXX, XXX−XXX
Terminal oxidation to methionine is also shown at low levels, this may be due to the exposure of material to oxidizing agents such as peracetic acid which is used to chemically disinfect the material during processing. In addition, EO sterilization may contribute to methionine PTM abundance. While disinfectant and sterilant exposure is necessary to ensure the safety of dECM materials for clinical use, there is evidence that such chemicals can alter methionine residues and it is possible that modifications caused by EO may have reduced the number of proteins identified. However, a study comparing dECM materials before and after terminal sterilization by EO demonstrated that there was no significant changes in relation to biological properties of the material.

Alternate solubilization methods were employed to explore the less abundant proteins of OFM. It is clear from our understanding of soft tissue repair that a large number of noncollagenous ECM proteins are involved in complex processes such as tissue homeostasis, wound healing, and tissue regeneration. In order to describe the complete inventory of OFM proteins, it is necessary to use methods that enrich smaller less abundant proteins than collagens. In this approach (Methods C and E), size exclusion chromatography was used to enrich for small proteins such as growth factors. OFM is known to be composed of major ECM structural proteins such as collagen and elastin; these proteins are generally large insoluble structures that may mask the identification of smaller soluble proteins such as secreted factors. A sample of CTGF was used to identify the likely fractions that would contain smaller proteins after SEC, this lead to the pooling strategy of combining fractions based on the elution of CTGF in order to target small proteins. CTGF was used because it is a well-known ECM-associated growth factor; however, many other substitutes for a molecular weight marker could be used.

The signal intensity of identified proteins can be used the estimate their abundance within the sample; however, since this method enriches soluble proteins over nonsoluble collagens it cannot be taken as an accurate quantification of the material as a whole but instead provides a relative abundance. The intention here was to describe non-collagenous ECM proteins that are sometime referred to as secondary molecules in the ECM.

This enrichment strategy shifted the abundance of proteins by their matrisome category, as shown in Figure 3 protein samples from the initial quantitative MS showed a high proportion of proteins are collagens (90%), while saline and GndHCl samples have a lower proportion of collagens (~30%) and a much higher portion of glycoproteins, secreted factors, and “other” proteins. Identification of proteins by MS from such a complex sample can become challenging due to ion interference; consequently, fractioning and pooling proteins by SEC has led to the identification of a much larger number of proteins, as shown in Table 2 and Figure 4.

There were small differences in the data sets generated from Method C and E (saline versus GndHCl extraction). For instance, a larger proportion of ECM regulators are identified by Method E while a larger proportion of secreted factors are identified in the saline extract (Method C).

The collagens are the most abundant proteins in the matrisome of most ECMs, the in silico database includes 45 genes in the human matrisome. As shown in Figure 5, 23 peptides were identified from 13 different types of collagens, the most abundant in these samples are collagen I, III, V and VI. When compared to published collagen composition of other ECMs, this demonstrates that OFM retains a diverse array of native ECM collagens, similar to the profile of collagen demonstrated in urinary bladder matrix (UBM) tissue, and with a greater diversity than highly processed collagen from connective tissue (e.g., rat tail collagen).

Collagens make up the fundamental building blocks of the ECM, but it is also recognized that they contain active sites and binding sites that are important during tissue remodeling. Collagens can be designated as “fibril forming”, “fibril associated”, “network forming” collagens and “membrane associated”. As shown in Figure 5, OFM retains a diverse array of collagens including fibril forming (I, II, III, and V), fibril associated and network forming. Previously, collagens in OFM have been characterized, to include collagens I and III and collagen IV, a major component of the basement membrane. In this study, these and many additional collagens have been identified as part of the OFM proteomic inventory. Interestingly, collagen V was identified as an abundant protein in this category (see Figure 4). Collagen V is known to be essential in the fibrillation of types I and II collagen and is found at the interface between dermal and epidermal tissue and also in placental ECM.

The glycoproteins represent a much larger array of proteins that are less abundant than collagen in tissue ECM. Like collagens, some glycoproteins such as elastin, fibronectins, and laminins can form fibrous structures adding to the structural part of the ECM, while others form a hydrogel like structure in the interstitial space. Previously, OFM was known to contain elastin, laminin, fibronectin. This study revealed a large number of less abundant glycoproteins present in OFM. As shown in Table 2, 59/195 known glycoproteins were identified. This number is greater than the number of glycoproteins identified in UBM (26/159) and rat tail collagen (5/159).

Major ECM glycoproteins are dynamic molecules than interact with both cells within tissue and binding sites of structural ECM proteins. Some interesting findings in this category are fibulin, basement membrane binding glycoproteins, and SPARC (osteonectin), which is known to be highly expressed in tissue with high turnover such and intestinal epithelium. Thrombospondins, emilins, and tenascins are classed as “matricellular” proteins, because of their diverse unknown functions in the ECM. Tenascin in particular is known to alter many biological processes including tissue remodeling.

The basement membrane component laminin has been previously identified in OFM; however, other basement membrane proteins nidogens 1 and 2 were identified using this approach. The vascular ECM proteins fibrinogen, vitronectin and WIF were also identified, which is somewhat expected as OFM contains vascular channels that are visible to the naked eye. Vitronectin is known to have an important role in the control of the clotting response.

Another important function of the ECM is the binding and storage of growth factors. Latent TGFβ binding proteins 1, 2, and 4 as well as insulin like growth factor binding protein 3 were identified in this study. The ECM stores growth factors that interact with cells when they are required, for instance, during tissue remodeling. Growth factor binding proteins along with GAGs and proteoglycans are responsible for storing growth factors and regulating their receptor access to surrounding cells.
Proteoglycans are glycoproteins that contain one or more GAG side chains. These are classified in terms of the GAG side chains and repeating leucine rich repeat (LRR) or LINK domains. Basement membrane proteoglycans identified in this study include perlecan and serglycin. Perlecan, the most abundant proteoglycan in OFM (see Figure 5) is a structural and function protein that binds and cross-links other ECM proteins, and maintains vascular homeostasis by inhibiting smooth muscle cell proliferation.41–63

LRR proteoglycans including decorin, lumican, and podocan are classified by their GAG side chains and LRR domain.40,42,64–66 Some of these have important structural roles involving the spacing and organization of collagen I and other fibrous structures. In addition, these proteoglycans have multiple functional biological roles in tissue homeostasis and remodeling. Versican contains a LINK domain rather than an LRR and is known to have various cell modulating functions including cell migration, adhesion a proliferation.

Several ECM proteins, especially of the collagen family are known to contain matrincines or matrixcryptic proteins, which are bioactive fragments that are release when the matrix is remodelled by the activity of proteases.1,57,67 These bioactive fragments often have a separate biological activity to their parent proteins via the association of hidden ligands with integrins, and can be involved in a large number of processes such as growth factor binding, cell migration, promotion of inhibition of angiogenesis.68,69 In Figure 5, proteins that were identified in OFM that have known matrikines or matrixcryptines associated are indicated by an asterisk (*) according to reviews on matrikines and cryptic peptides.41

Generally the least abundant matrisome proteins are the ECM regulators and secreted factors, including growth factors, and protease inhibitors.7 Growth factor proteins are often small, labile molecules. These proteins are key to dynamic reciprocity between cells and the ECM, and are constantly expressed by cells within the ECM to control the construction and destruction of the ECM.1,57,70 These signaling chemokines are beneficial components during tissue remodeling but are labile and short-lived, making them difficult to retain in dECM biomaterials.

Growth factors are continually expressed by cells during paracrine signaling; however, some are sequestered by ECM proteins to induce a quick response when needed.52 Upon release, growth factors are activated to provide a transient signal to cells within the matrix. The ECM in turn controls access of growth factors to cells to maintain the balance of tissue remodeling and repair. Growth factors such as PDGF and TGFβ are released upon damage to the ECM, and produce a signal to activate wound healing cascades such as fibroblast recruitment, proliferation, and collagen synthesis.60

Previous studies have demonstrated the presence of FGF2 in OFM.52 Using this approach additional growth factors were identified as listed in Figure 5. Pooled extracts were also subject to a targeted approach to identify TIMPs and growth factors in this sample. This led to the identification of FGF7, EGF, and several BMPs (bone morphogenic proteins).

There are a large number of growth factors in the in silico matrisome; however, their identification by MS is challenging due to their low abundance and the complexity of samples. There are some examples of growth factors that have been identified by MS in native tissue (e.g., FGF in liver and colon).30 More often growth factors are identified using a direct approach such as ELISA. For instance, growth factors have been identified in UBM by ELISA,18 but they do not appear in the data set generated by MS.71 Similarly, a direct approach was used to identify growth factors in porcine small intestinal submucosa (SIS) and decellularized human amnion membrane tissue (dHAM).72–75

While growth factors have previously been identified in dECM biomaterials using targeted approaches such as ELISA or Western blot, to our knowledge this is the first example of a commercial dECM biomaterial growth factor identification using a MS based proteomic approach. A number of studies have identified growth factors in lab scale decellularized tissues; however, for clinical safety commercial biomaterials are generally processed more thoroughly to remove DNA and other DAMPs (damage-associated molecular patterns) and are subject to terminal sterilization processes.

ECM regulators are mostly enzymes and inhibitors that are responsible for the maintenance of the ECM in resting and remodeling tissue.7 These proteins are expressed by surrounding cells to make or break the matrix as required. A number of enzymes involved in collagen and other protein cross-linking were identified in this study, such as lysyl oxidase homologue. Proteases responsible for processing ECM proteins were also identified, including two ADAMTS proteins and Cathepsin D, a well-known enzyme with numerous physiological functions including the activation and degradation of polypeptide hormones and growth factors.76

Enzymes that modulate the matrix are tightly controlled by inhibitors such as TIMPs. TIMPs are an integral part of the ECM responsible for controlling the breakdown of damaged tissue, for instance, during proteolysis of damaged tissue resulting from an aberrant immune response.77 It is believed that the presence of TIMPs in a biomaterial alleviate the uncontrolled proteolytic activity of immune cells in chronic wounds.78,79 This study identified TIMP4, as well as, three serpins. Serpins control a number of biological processes indulging coagulation and inflammation.80 This identification of protease inhibitors reiterates previous work by Negron et al. describing the ability of OFM to control proteases such as neutrophil elastase and matrix metalloproteases (MMPs).77

While a large number of unknown proteins have been identified in this list, there are several notable limitations to this approach. First, as is shown in Figure 2, certain proteins were identified by targeted SRM approach only and not the global LC−MS approach and vice versa. Importantly, the presence of collagen III in OFM has been demonstrated using other methods;21 however, the peptides used in this targeted approach did not identify this protein.

Second, it is possible that the process of decellularization and sample preparation gave rise to unknown PTMs that were not considered during our search; this may have affected protein identification using this method.

Third, although the abundance of proteins is described using their signal intensity, this is only a "relative abundance" based on the extraction, which has enriched soluble proteins over insoluble collagens. This approach allows more proteins to be identified but negates the ability to quantify these proteins as a proportion of the material as a whole.

Finally, information about the native proteome is derived from other studies using human and mouse tissue; although collagens are highly preserved between mammals, certain assumptions have been made regarding the conservation of other ECM genes between human and ovine genome.
CONCLUSIONS

The diversity of the matrisome is vital for numerous processes in living tissue, including homeostasis, wound healing, growth, and development. The sheer number of proteins and differences in their abundance in different tissue types demonstrates how complex a system is being orchestrated between cells and the many proteins of the ECM. An extensive proteome demonstrates how complex a system is being orchestrated between different methods of sample preparation along with global or targeted protein identification will lead to the discovery of different ECM protein data sets; no single method can capture the total complexity of tissue-derived biomaterials. Biomaterials that recapitulate the complex molecular diversity of the native ECM are therefore highly suited for clinical applications to support repair and regeneration of tissue. In vivo, communication between cells and the ECM enables a number of mechanisms required for tissue remodeling, some of which are not yet fully understood. This work highlights the degree and diversity of molecular cues present in natural biomaterials and may lead to a better understanding of the dECM mediated remodeling.

REFERENCES

(4) Badylak, S. F. The extracellular matrix as a biologic scaffold material. Biomaterials 2007, 28 (25), 3587−93.
(10) Knothe Tate, M. L. Top down and bottom up engineering of bone. J. Biomech. 2011, 44 (2), 304−312.
(12) Burns, J. S.; Kristiansen, M.; Kristensen, L. P.; Larsen, K. H.; Nielsen, M. O.; Christiansen, H.; Nehlin, J.; Andersen, J. S.; Kassem, M. Decellularized matrix from tumorigenic human mesenchymal stem cells promotes neovascularization with galectin-1 dependent endothelial interaction. PLoS One 2011, 6 (7), No. e21888.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00908.

Inventory of proteins identified in OFM by all methods described in this manuscript (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: barnaby.may@arobio.com.

ORCID

Sandi G. Dempsey: 0000-0001-9855-018X

Notes

The authors declare the following competing financial interest(s): BCHM and CHM are shareholders of Aroa Biosurgery Limited. RH and KH are shareholders in Omix technologies LLC.

ACKNOWLEDGMENTS

This research was undertaken by funding provided by Aroa Biosurgery and Callaghan Innovation Limited (New Zealand) Growth Grant (MSMA1402).

ABBREVIATIONS

OFM, ovine forestomach matrix; dECM, decellularized extracellular matrix; GAG, glycosaminoglycans; CTGF, connective tissue growth factor; EO, ethylene oxide; TIMPs, tissue inhibitors of metalloproteinases; UBM, urinary bladder matrix; LRR, leucine rich repeat; SIS, small intestinal submucosa; dHAM, decellularized human amnion membrane tissue.


