

# Characterization of Neopeptides in Equine Articular Cartilage Degradation

Mandy Jayne Peffers,<sup>1</sup> David James Thornton,<sup>2</sup> Peter David Clegg<sup>1</sup>

<sup>1</sup>Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Leahurst, Chester High Road, Neston, Wirral CH64 7TE, UK, <sup>2</sup>Wellcome Trust Centre for Cell Matrix Research, Faculty of Life Sciences, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

Received 27 February 2015; accepted 4 June 2015

Published online 7 July 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jor.22963

**ABSTRACT:** Osteoarthritis is characterized by a loss of extracellular matrix that leads to cartilage degradation and joint space narrowing. Specific proteases, including the aggrecanases ADAMTS-4 and matrix metalloproteinase 3, are important in initiating and promoting cartilage degradation in osteoarthritis. This study investigated protease-specific and disease-specific cleavage patterns of particular extracellular matrix proteins by comparing new peptide fragments, neopeptides, in specific exogenous protease-driven digestion of a crude cartilage proteoglycan extract and an in-vitro model of early osteoarthritis. Additionally, equine cartilage explants were treated with interleukin-1 and the media collected. Proteolytic cleavage products following trypsin digestion were then identified using tandem mass spectrometry. Complete sequences of proteolytically cleaved neopeptides were determined for the major cartilage proteoglycans aggrecan, biglycan, decorin, fibromodulin plus cartilage oligomeric matrix protein. The generation of neopeptides varied with enzyme specificity; however, some peptides were common to all samples. Previous known and novel cleavage sites were identified. The identification of novel peptide fragments provides a platform for the development of antibodies that could assist in the identification of biomarkers for osteoarthritis (OA), as well as the identification of basic biochemical processes underlying OA. © 2015 The Authors. *Journal of Orthopaedic Research* Published by Wiley Periodicals, Inc. on behalf of Orthopaedic Research Society. *J Orthop Res* 34:106–120, 2016.

**Keywords:** osteoarthritis; neopeptides; cartilage; equine

The unique load-bearing properties of articular cartilage are dependent upon its structural composition and organization, particularly the interactions between collagens and proteoglycans of the extracellular matrix (ECM).<sup>1</sup> In normal physiology, these matrix macromolecules are turned over by chondrocytes embedded within the cartilage.

Progressive degeneration of articular cartilage, including proteolysis-driven degradation, leads to joint pain and dysfunction that is clinically identified as osteoarthritis (OA). Under normal circumstances, there is an equilibrium between matrix deposition and degradation; however, this equilibrium is disrupted in OA, leading to the excessive degradation of matrix and progressive loss of important matrix components, such as collagens and proteoglycans.<sup>2,3</sup> Furthermore, intact and fragmented ECM peptides, produced following degradation, affect chondrocyte function through integrin receptor signaling.<sup>4</sup>

Studies have identified peptide fragments (neopeptides) of the ECM constituents in cartilage metalloprotease digests,<sup>5</sup> tendon ageing<sup>6</sup> and disease,<sup>6,7</sup> and OA.<sup>8</sup> In cytokine-stimulated models of OA, the breakdown of cartilage is stimulated by a series of proteolytic enzymes through the upregulation of metalloproteinase.<sup>9</sup> Initially, aggrecan is fragmented and released

from cartilage in OA, followed by other molecules, such as cartilage oligomeric matrix protein (COMP), fibromodulin, and collagens.<sup>10</sup> Furthermore, these neopeptides can be used as biochemical markers.

Members of the matrix metalloproteinases (MMPs) and disintegrin and metalloproteinase with thrombospondin type I motifs (ADAMTS) families of enzymes are important in cartilage matrix degradation in OA.<sup>11</sup> MMP-3 is one of the most highly expressed proteases in cartilage capable of degrading proteoglycans, including aggrecan,<sup>5</sup> as well as activating procollagenases. ADAMTS-4 is an important enzyme in the pathogenesis of OA as demonstrated by its high aggrecanase activity in OA cartilage, and localized expression in the areas of aggrecan degradation.<sup>12</sup> Both ADAMTSs and MMPs are involved in the cleavage of other ECM proteoglycans, including aggrecan,<sup>13</sup> biglycan<sup>14,15</sup> and decorin,<sup>16</sup> fibromodulin,<sup>17</sup> and COMP.<sup>18</sup>

Diagnosis of OA at an early stage is difficult as it is based on radiographic changes and clinical symptoms that occur relatively late in disease. However, the advent of advanced proteomic techniques has enabled the identification and use of protein biomarkers to become established<sup>19</sup> that may aid in not only the monitoring of OA progression but also the effects of therapeutics for OA. Neopeptides provide potential biomarker candidates.

In this study, we use liquid chromatography tandem mass spectrometry (LC-MS/MS), de-novo sequencing, and database searching to enable an accurate and convenient workflow to identify neopeptides (in selected proteoglycans; important to cartilage biology) released by either specific protease treatment of cartilage proteoglycan extract or following cytokine stimulation of equine articular cartilage. One way to provide novel insights into development and treatment of OA is to obtain an overall understanding of cartilage degradation. This analysis was performed to increase the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Grant sponsor: Wellcome Trust Veterinary Integrated Research Fellowship; Grant number: WT088557MA; Grant sponsor: Liverpool Health Partners; Grant sponsor: MRC and Arthritis Research UK.

Correspondence to: Mandy Jayne Peffers (T: +44 151 794 2222; F: +0151 794 2000; E-mail: peffers@liv.ac.uk)

© 2015 The Authors. *Journal of Orthopaedic Research* Published by Wiley Periodicals, Inc. on behalf of Orthopaedic Research Society

knowledge of molecular events associated with cartilage degradation characteristic of OA, identify peptides that may be useful as biomarkers of cartilage disease, as well as to identify novel neopeptides that could be further validated in targeted future studies.

## EXPERIMENTAL PROCEDURES

All chemicals were supplied by Sigma–Aldrich, Dorset, UK unless otherwise stated.

### Cartilage Sampling and Procurement

Samples were collected as a by-product of the agricultural industry. Specifically, the Animal (Scientific Procedures) Act 1986, Schedule 2, does not define collection from these sources as scientific procedures. Ethical approval was, therefore, not required for this study. Full-thickness cartilage from the entire surface of macroscopically normal metacarpophalangeal joints of three independent horses for each study (explant; mean age  $7 \pm 1$  years, proteoglycan isolation  $8.3 \pm 0.6$  years) was collected from an abattoir. Macroscopic scoring was undertaken using a macroscopic grading system as described previously.<sup>20</sup>

### Normal Equine Cartilage IL-1 $\beta$ -Treated Explant Studies

Cartilage explant studies were undertaken as previously described.<sup>21</sup> The media was exchanged 48 h after initiation of treatment, and cultures harvested after 96 h. The 48- and 96-h supernatant samples were pooled, thus representing the total secretome over 96 h.

### Proteoglycan-Enriched Fraction Isolation From Cartilage

Proteoglycan was extracted from cartilage pooled from three donors using 4 M guanidinium chloride as previously described.<sup>6</sup> The soluble fraction was removed following centrifugation for 15 min at 13,000g at 4°C and dialyzed in a 14,000-kD cut-off membrane (Spectrapor, Breda, the Netherlands) for 24 h at 4°C against 0.1 M sodium acetate, pH 6.0 in the presence of protease inhibitors. The extract was centrifuged for 15 min at 13000g at 4°C. The supernatant was fractionated in an associative cesium chloride (CsCl) density gradient (starting density 1.5 g/ml) for 60 h at 100,000g in an ultracentrifuge (Beckman 50Ti, Galloway, Ireland). The tube was fractionated into quarters, A1–A4. The combined A1–A2 fractions, identified in previous studies as being enriched for aggrecan and small leucine-rich proteoglycans,<sup>22</sup> were retained for protease digestion and dialyzed against 0.1 M sodium acetate for 48 h at 4°C and then against ultrapure water for 36 h at 4°C. The samples were then lyophilized. An aliquot of each fraction was assessed for protein content using optical density of 280 nm with a Nanodrop ND-100 spectrophotometer (Labtech, East Sussex, UK). To validate, the A1–A2 fractions were enriched for proteoglycans, glycosaminoglycan (GAG) analysis of the A1–A4 fractions was undertaken using a 1,9-dimethyl-methylene blue (DMMB) dye-binding microwell spectrophotometric assay.<sup>23</sup>

### Protease Digestion of the Proteoglycan Extract

Aliquots of the A1–A2 extract were digested in protease digestion buffer (50 mM Tris HCl, 100 mM NaCl, 10 mM CaCl<sub>2</sub>, pH 7.5) with either 0.05 nmol human recombinant MMP-3 catalytic domain (Calbiochem, La Jolla, CA) for 20 h at 37°C or with 0.014 nmol truncated human recombinant ADAMTS-4 (Calbiochem) for 7 h at 37°C. Times were chosen based on preliminary studies (data not shown). A control for each protease was incubated under the same conditions in the presence of the recombinant protein formulation buffer. The enzymatic digestion reactions were stopped by addition of EDTA.

### Deglycosylation of the Proteoglycan Extract and Immunoblot Analysis

A1–A2 extracts crude proteoglycan extracts (CPE) before and after protease digestion were deglycosylated as previously described.<sup>13</sup> Deglycosylated samples of the CPE, equivalent to 5  $\mu$ g GAG were applied to the nitrocellulose membrane. Samples were prepared for immunoblotting as described.<sup>24</sup> The membrane was probed overnight at 4°C with the following antibodies in TBS-T containing 5% milk: mouse monoclonal to aggrecan ARGxx (BC-3) (Abcam, Cambridge, UK) (1:100 dilution), mouse monoclonal to aggrecan DIPEN (MD Bioproducts, Minneapolis, MN) (1:100 dilution), and rabbit polyclonal to aggrecan (Abcam) (1:1000). The following secondary peroxidase conjugated antibodies were used: goat anti-mouse IgG and goat anti-rabbit IgG both at 1:1000 dilution (Abcam).

### In-Solution Tryptic Digestion

Samples of cartilage supernatant from the explant experiments and proteoglycan extract from the protease digestion experiments were trypsin digested as described previously.<sup>6</sup> CPE samples were desalted and purified using C<sub>18</sub> resin in the form of a ZipTip<sup>®</sup> (Merck Millipore, Billerica, MA).

### LC-MS/MS Analysis

LC-MS/MS analysis was performed using NanoAcquity<sup>™</sup> Ultraperformance LC (Waters, Manchester, UK) on line to an LTQ-Orbitrap Velos (Thermo-Fisher Scientific, Hemel Hempstead, UK). Aliquots of tryptic peptides equivalent to 250 ng were loaded as previously described.<sup>21</sup>

### Neopeptide Identification

For neopeptide identification, raw spectra were converted to mascot generated files (mgf) using Proteome Discoverer software (Thermo, Hemel Hempstead, UK). The resulting mgf files were searched against the Unihorse database using an in-house Mascot<sup>25</sup> server (Matrix Science, London, UK). Search parameters used were as follows: enzyme; none, peptide mass tolerances 10 ppm, fragment mass tolerance of 0.6 Da, 1+, 2+, and 3+ ions, missed cleavages; 1, and instrument type ESI-TRAP. Modifications included were as

follows: fixed; carbamidomethyl cysteine and variable; oxidation of methionine. The probability that a match was correct ( $p < 0.05$ ) was determined using the Mascot-derived ion score, where  $p$  was the probability that the observed match was a random event. For reasons of economy and to have confidence in our analysis, we only included neopeptides in the results if they were present in treated samples exclusively as identified by Mascot. Patterns of fragmentation were determined for aggrecan, biglycan, decorin, fibromodulin, and COMP.

In addition, raw data files were loaded into PEAKS<sup>®</sup> Studio 6.0 (Bioinformatics Solutions, Inc., Waterloo, Canada) and de-novo sequencing and protein identification performed. Estimate FDR function was used in order to create a “decoy fusion” database based on the Ensembl database for horse (*Equus caballus*; EquCab2.56.pep, (ftp://ftp.ensembl.org/pub/current/fasta/equus\_caballus/pep/)). Results generated using PEAKS<sup>®</sup> Studio was manually curated against the Mascot search engine results. A 10lgP score of  $>20$  was considered as significant. Mass spectrometry data is available in PRIDE are available in the PRIDE database (<http://www.ebi.ac.uk/pride/>) at the European Bioinformatics Institute under accession numbers 19447–19479.

## RESULTS

### Production of a Proteoglycan-Enriched Fraction by Cesium Chloride Density Gradient-Ultracentrifugation

In this study, we were interested in identifying potential cleavage sites in cartilage proteoglycans and COMP. Therefore, the CPE, which was also found to be abundant in COMP, was purified using cesium chloride density gradient centrifugation. Proteoglycans were found, as validated by the GAG assay results at the expected density of 1.3–1.5 g/ml (Fig. 1). Proteins were predominantly in the A4 fraction as determined by protein absorbance (data not shown). The proteoglycan-rich A1–A2 fractions were then pooled for further work.

### Human Recombinant Proteases Were Validated for the Use in Equine Cartilage Proteoglycan-Enriched Fraction Digestion

As equine recombinant proteases relevant to this study were not freely available, human recombinant proteases were used. Crude proteoglycan extracts were analyzed before (control) and after MMP-3 or ADAMTS-4 digestion. Immunoblotting was undertaken with neopeptide antibodies. Neopeptides are defined as part of a molecule that is the target of an immune response. In this case, the target is a peptide fragment. Immunoblotting using anti-ARGxx antibody (ADAMTS-4 derived neopeptide), anti-DIPEN antibody (MMP-3 derived neopeptide), and anti-aggrecan revealed intact aggrecan, as well as degradation products consistent with the activity of human recombinant proteases on equine aggrecan. ADAMTS-4 proteolysis (but as expected, not MMP-3) produced a product identified by

the aggrecanase derived antibody ARGxx (Fig. 2A) and MMP-3 digestion produced a product identified by the DIPEN antibody (Fig. 2B).

### Identification of Neopeptides in Protease-Treated Proteoglycan-Enriched Fraction

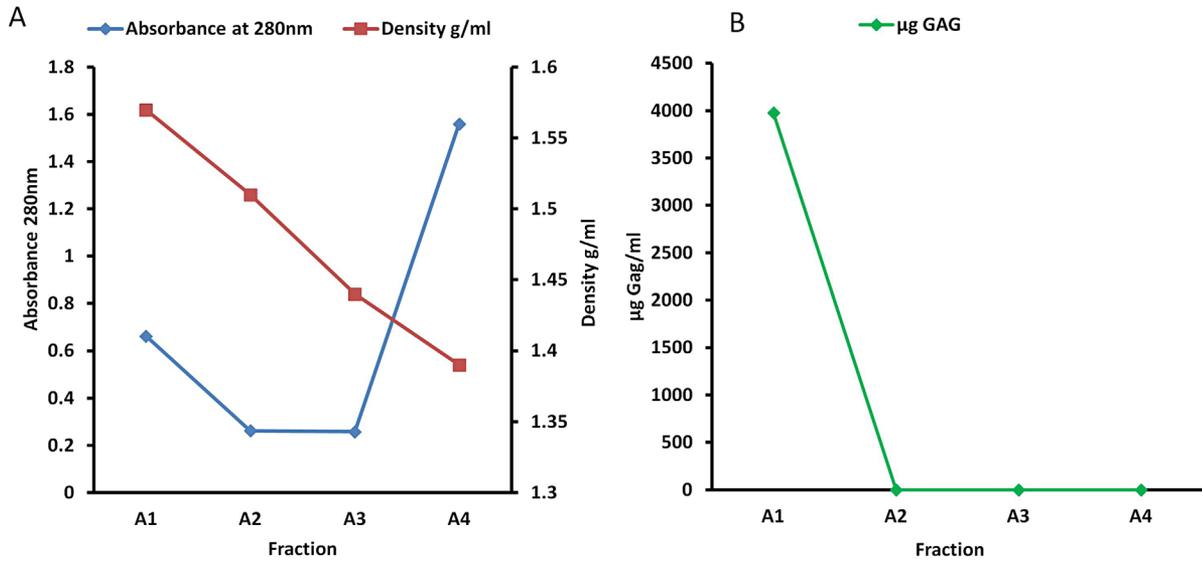
Mascot identified 87 and 84 significant proteins with more than one unique peptide for the ADAMTS-4 and MMP-3 digests, respectively. The proteoglycans (aggrecan, biglycan, decorin, and fibromodulin) investigated in this study and COMP were in the top 10 proteins identified in all samples. Supplementary Table S1 provides detailed information on the identification of peptides mapped to each protein and corresponding Mascot scores. The numbers of neopeptides that were generated are demonstrated in Figure 3. Table 1 indicates the number of unique neopeptides identified for each protease digestion. Table 2A and B (ADAMTS-4) and Table 3A and B (MMP-3) demonstrate sequences of neopeptides. The number of times each neopeptide was identified varied from once (76% and 74% in ADAMTS-4 and MMP-3 digestion, respectively) to 33 times within a single experiment for the MMP-3 derived biglycan neopeptide <sup>152</sup>NHLVEIPPNLPS<sup>164</sup>.

### Identification of Neopeptides Following IL-1 $\beta$ Stimulation of Articular Cartilage Explants

Each donor was analysed separately. In control media, we identified  $141 \pm 4.7$  and in the IL-1 $\beta$ -treated media  $226 \pm 17.6$  proteins. Supplementary Table S2 contains complete lists of all peptides identified. A catalog of IL-1 $\beta$ -generated neopeptides (defined here as those identified in at least two treated donors but no controls) from each protein of interest was identified (Table 4A and B). Table 5 shows the neopeptides identified in multiple conditions in this study or those previously identified.

## DISCUSSION

We have performed a proteomic analysis of crude proteoglycan extract identifying protease-related neopeptides within the cartilage proteoglycan-enriched fraction and compared these to neopeptides produced in an IL-1 $\beta$  driven in-vitro early inflammatory model of OA. OA is a syndrome, and it is likely that not all forms of OA are directly related to IL-1. Recent work has disputed a substantial role for IL-1 in OA. There is no change in the concentration of IL-1 $\beta$  in OA synovial fluid,<sup>26</sup> the use of IL-1 $\beta$  receptor antagonist has had variable results,<sup>27,28</sup> the role in murine OA is also disputed.<sup>29,30</sup> In addition, most in-vitro experiments use super physiological concentrations of IL-1 $\beta$  (synovial fluid concentrations are 10 pg/ml). Thus, while a role can be argued for IL-1 $\beta$  in OA achieving, clinical relevance may require additional factors in in-vitro studies. However, we feel that IL-1 is still relevant cytokine to study in relation to many pathogenesis's of OA.<sup>31</sup> Some of the neopeptides were generated by both proteases and in the IL-1 $\beta$ -treated



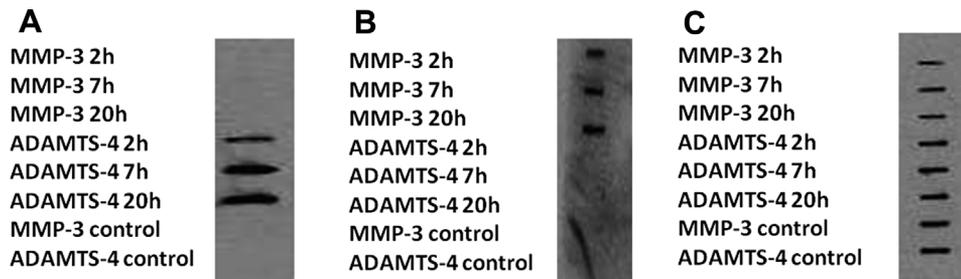
**Figure 1.** Guanidinium chloride extraction and CsCl centrifugation produced a proteoglycan-enriched fraction of three pooled samples of equine cartilage. Solid CsCl was added to extracted soluble cartilage proteins at a starting density of 1.5 g/ml. After centrifugation, the tubes were fractionated into four equal fractions and the density of each fraction was measured. Fractions were then measured for (A) protein absorbance at 280 nm, and assayed to determine (B) GAG concentrations of each fraction. Fractions A1 and A2 were then pooled for further work.

media, whereas others were unique for either a single protease or  $\text{II-1}\beta$  dependant. Furthermore, a number of previously identified neopeptides or cleavage sites pertaining to these were evident.

We hypothesise that a number of the neopeptides identified contain cleavage sites due to protease degradation, either MMP-3 or ADAMTS-4 or proteases downstream of  $\text{II-1}\beta$ . These proteases have been implicated in the pathogenesis of OA.<sup>3,14,31–33</sup> We were interested in important cartilage proteins with diverse functions, including key regulators of collagen fiber assembly; the small leucine-rich proteoglycans; biglycan, decorin, fibromodulin, and COMP. In addition, we studied aggrecan neopeptides as this is the major proteoglycan in cartilage, enabling the tissue to resist compressive loads. Therefore, rather than extracting proteins from whole cartilage, a crude proteoglycan

extract was prepared. Classic methodologies, previously used to identify cartilage fragments in aggrecan,<sup>34</sup> were used in order to remove proteins of non-interest, thus reducing sample complexity for downstream LC-MS/MS. The A1–A2 fractions contained the proteoglycans of interest, including aggrecan as determined by immunoblotting. Subsequent analysis at the protein level demonstrated that the abundant proteins in the crude proteoglycan extract were the ECM proteins in which we were interested.

We used equine samples in this study as the horse is an athletic animal and is considered an excellent animal model for human joint diseases due to extensive knowledge of its pathogenesis and clinical experience of the disease.<sup>35</sup> Equine tissue was readily obtained, enabling collection of cartilage samples from macroscopically normal horses.



**Figure 2.** A slot blot of MMP-3 and ADAMTS-4 digests of cartilage proteoglycan-enriched extracts using anti-neoepitope antibodies reveals that human recombinant proteins are active against equine aggrecan. Proteoglycan-enriched fractions were extracted from pooled equine cartilage of three donors and analyzed both before and after ADAMTS-4 or MMP-3 digestion using the following antibodies: (A) anti-ARGxx, (B) anti-DIPEN, and (C) Anti-aggrecan. Prior to ADAMTS-4/MMP-3 digestion, only bands identifying intact aggrecan were seen. Following ADAMTS-4 digestion at all incubation times, ADAMTS-4-derived degradation products were evident. Similar results were identified following MMP-3 digestion.



**Table 2.** Neopeptides Identified Following A. ADAMTS-4 Digestion of Crude Proteoglycan Was Identified With LC-MS/MS Using Mascot

A							
Protein	Amino Acid Before Peptide Residue	Peptide Sequence	Amino Acid Residue After	Significant MASCOT Identified Peptides; $p < 0.05$	Significant PEAKS 10lgP	Position	
Aggrecan	R	LATTGQLY <sup>#</sup>	L	*	*	280–287	
	R	VSLPNYPAIPTDATLELQN <sup>#</sup>	L		*	105–122	
	R	LATTGQLYLAW <sup>#</sup>	Q		*	280–290	
	V	EDISGFPSGGE <sup>#</sup>	V		*	989–998	
	R	WSDGHSLQFEN <sup>#</sup>	W		*	2275–2285	
	P	WATEVPSASEKPSPE	E			787–802	
	R	VSLPNYPAIPTDATLE	L			104–119	
	R	YDAICYTGEDFVDIPENFFA	V			344–363	
	D	LSGTSGRAD	V			1762–1771	
	Q	EAGEGPSGILE	L			1838–1848	
	F	RGQPSEGSVSL	P			834–845	
	G	EGPSGILELSGAHSG	A			1841–1855	
	I	HDLVSSAMSGSE	P			1698–1710	
	G	FSGTTSGIHDLVSSAMSG	S			1702–1716	
Biglycan	A	SGVEDLGGLPSGGEIHLEPTASGV	E			967–988	
	K	NHLVEIPPNLPS <sup>#*1</sup>	S	*	*	152–163	
	K	LLQVVYLH <sup>#</sup>	T	*	*	305–312	
	N	GISLFNNPVPY <sup>#</sup>	W	*	*	338–348	
	N	CIEMGGNPLENSGFQPGAFDGLK <sup>*2*3</sup>	L	*		191–213	
	K	DLPETLNELHLDHN	K		*	230–243	
	F	NNPVPYWEVQPATF	R		*	343–356	
	A	IELEDLLR	Y		*	248–255	
	F	NNPVPYWEVQPA	T		*	343–354	
	M	CPFGCHC	H			67–73	
	D	SLTPTFSAMCPF	G		*	58–69	
	K	DLPETLNE	L			230–237	
	COMP	R	AFQTVVLDPEGDAQIDPN <sup>#</sup>	W	*	*	531–548
		K	QVCTDIDECETGQHNCVFN <sup>#*3,7</sup>	S	*	*	174–192
N		WVVLNQGMEIVQTM <sup>#</sup>	N	*	*	549–562	
Q		CAPGSCFPGVAC <sup>#*4,7</sup>	S	*	*	89–100	
R		WFLQHRPQVG <sup>#</sup>	I	*	*	682–692	
K		NTVMECDACGMQP <sup>*5</sup>	A		*	63–75	
Decorin	R	CEACPPGYSGPHEGVGMAT <sup>*6,7</sup>	A		*	149–168	
	R	IHEVLDLEPLGPVCPF <sup>#</sup>	R	*	*	42–57	
	K	NLHALILVN	N			107–117	
	K	ASYSGVSLFSNPVQYWEIQPS	T		*	323–343	
	K	YIQVVYLH	N		*	294–301	

B

B						
Protein	Amino Acid Before Peptide Residue	Peptide Sequence	Amino Acid Residue After	Significant MASCOT Identified Peptides; $p < 0.05$	Significant PEAKS 10lgP	Position
Decorin	Q	MIVVELGTNPLKSSG <sup>*3</sup>	I			177–191
Fibromodulin	R	DCPQECDCPPNFPTAMY <sup>#</sup>	C	*	*	75–91
	R	LSHNSLTNN <sup>#</sup>	G		*	275–283
	R	IPPVNTNLEN <sup>#</sup>	L		*	308–317
	R	INEFSISSFC	T		*	325–334
	R	KVPDGLPSALEQLYLEHNNVY	S		*	237–257

TABLE 2  
Continued

B

Protein	Amino Acid Before Peptide Residue	Peptide Sequence	Amino Acid Residue After	Significant MASCOT Identified Peptides; $p < 0.05$	Significant PEAKS 10lgP	Position
	R	ELHLDHN* <sup>1</sup>	<b>Q</b>		*	180–186
	R	KVPDGLPSALEQLYLEHNNVYSVPDSY	<b>F</b>		*	237–263

In the peptide sequence column, \* denotes neopeptides produced at known cleavages sites to specific proteases at the N or C terminal end and # denotes neopeptide that have been identified following protease digestion with both ADAMTS-4 and MMP-3. A significant PEAKS<sup>®</sup> derived 10lgP value is >20 that equates to  $p$  value <0.05. Potential cleavage sites are marked in bold. Position refers to the amino acid of the peptide sequence within the protein.\*<sup>1</sup> Ref.<sup>8</sup>;<sup>2</sup> Ref.<sup>15</sup>; \*<sup>3</sup> Ref.<sup>45</sup>; \*<sup>4</sup>Ref.<sup>7</sup>; \*<sup>5</sup>Ref.<sup>6</sup>; \*<sup>6</sup>Ref.<sup>46</sup>; \*<sup>7</sup>Ref.<sup>49</sup>.

the biglycan peptide <sup>337</sup>GISLFNNPVPY and the COMP neopeptide <sup>89</sup>CAPGSCFPGVAC. This could indicate that under some conditions, the two proteinases have the same specificity. Although an alternative hypothesis would be that this is due to the activation of ADAMTS4 by MMP3-mediated C-terminal processing of ADAMTS4, such as previously suggested for MMP17.<sup>38</sup>

Of the proteins investigated, COMP had the most neopeptides identified following Il-1 $\beta$  stimulation of cartilage explants. This could be due to a number of factors. COMP, a pentameric protein has a relatively even distribution of arginines and lysines within its sequence, making it an ideal for tryptic digestion producing tryptic (or semi- or none-tryptic in the case of neopeptides) masses of optimum mass range for detection. While from our previous IL-1 $\beta$  explant studies, in molar concentration terms, it is the most abundant protein.<sup>21</sup> It also has few post-translational modifications, such as glycosylations. These peptides would not be identified with the protein identification software used in this study. Finally COMP has been investigated as a musculoskeletal biomarker of OA<sup>39</sup> and tendon injury<sup>7</sup> in numerous studies and it is hypothesized that in arthritis, it is subject to extensive degradation accounting for the large number of neopeptides identified.

Our workflow used a mass spectrometer with high resolution and high mass/charge ( $m/z$ ) accuracy. For data processing, we used both the MASCOT search engine that uses of probabilistic search algorithms and the software PEAKS<sup>®</sup> enabling significantly improved sensitivity and through algorithms that enable sequencing de-novo.<sup>40</sup> As the aim of our approach was to provide a “first pass” list of peptides for further investigation, it was important that peptide identification results were statistically validated to avoid false positives. Therefore, in the PEAKS<sup>®</sup> analysis, we employed the decoy-fusion method that joins the decoy and target sequences of the same protein together as a “fused” sequence. The decoy fusion method avoids

some pitfalls in the standard target-decoy method, and is more conservative.<sup>40</sup>

We used the enzyme trypsin to produce high cleavage specificity peptides of suitable charge and sequence length for MS. Trypsin cleaves by the recognition of a target amino acid in a binding pocket. A negatively charged aspartate residue at the bottom of this pocket limits the amino acids to which this enzyme will recognize to arginine (R) and lysine (K) due to their long basic side chains with a few exceptions.<sup>41</sup> Thus, without further protease digestion, C-terminal R or K peptides are produced. Other possible fragmentations are produced by other proteases either from the environment (exogenous proteases that remove amino acids at the peptide terminals), the sample itself (for example following protease treatment) or following auto proteolysis (by intracellular enzymes and/or enzymes not normally active in the tissue) during the period between death and the addition of proteinase inhibitors. In order to minimize the latter, samples were collected by the removal of the leg distal to the carpus at the time of death and the cartilage removed within 2 h of death. Hence, semi-tryptic peptides or none tryptic peptides will possess a cleavage pattern independent of trypsin. These peptides are then subject to further fragmentation in the MS in order to identify the sequence of amino acids through the use of search engine specific algorithms. Thus, a workflow based on LC-MS/MS, de-novo sequencing and database searching provide an accurate and convenient method to identify peptide products released by either specific proteases or following cytokine stimulation.

A number of neopeptides pertaining to previously identified cleavage sites were evident. The neopeptide <sup>344</sup>YDAICYTGEDFVDIPEN corresponding to the DIPEN neopeptide<sup>42</sup> was demonstrated and confirmed with immunoblotting. However, the other major aggrecanase-derived cleavage sites (reviewed<sup>43</sup>) in aggrecan were not identified. This may be due to the size of the peptides produced being outside the mass range of the

**Table 3.** Neopeptides Identified Following MMP-3 Digestion of Crude Proteoglycan Were Identified With LC-MS/MS Using Mascot

Protein	Amino Acid Before	Peptide Sequence	Amino Acid After	Significant	Significant	Position
				MASCOT Identified Peptides; $p < 0.05$	PEAKS 10lgP	
Aggrecan	R	VSLPNYPAIPTDATLLELQN <sup>#</sup>	L	*	*	105–122
	R	LATTGQLY <sup>#</sup>	L	*	*	280–287
	R	VSLPNYPAIPTDATLE <sup>#</sup>	L	*	*	104–119
	R	LATTGQLYLAW <sup>#</sup>	Q	*	*	280–290
	V	EDISGFPSGGE <sup>#</sup>	V	*	*	989–998
	R	TYGVRPSSETYDVYC	Y	*	*	557–571
	R	ACLQNSAIATPEQL	Q	*	*	174–188
	R	ITCTDPASY	K	*	*	2419–2427
	R	ITCTDPAS	Y	*	*	2419–2426
	K	GTVACGDPPVVEH	A	*	*	2328–2340
	Y	LAWQSGMDC	S	*	*	288–297
	R	WSDGHSLQFENWR	P	*	*	2275–2287
	R	YDAICYTGEDFVDIPEN <sup>*8</sup>	F	*	*	344–380
	Y	QLPFTCK	K	*	*	2320–2326
	N	SAIATPEQL	Q	*	*	179–188
	R	PSSETYDVYC	Y	*	*	562–571
	R	TYGIRD	N	*	*	232–238
	K	GEWNDVPCNY	Q	*	*	2310–2319
	R	VSLPNYPAIPTD	A	*	*	104–115
	Biglycan	M	EGLETSASGAEDLSGL	P	*	*
K		LLQVVYLH <sup>#</sup>	S	*	*	305–312
K		NHLVEIPPNLPS <sup>#*1</sup>	L	*	*	152–164
N		GISLFNPNVPY <sup>#*3</sup>	W	*	*	338–348
K		VGVNDFCPVGFV	K	*	*	319–331
Y		NGISLFNPNVP <sup>*9</sup>	Y	*	*	337–347
K		VGVNDFCPVG	F	*	*	319–28
K		DLPETLNE	L	*	*	230–237
R		NMNCIEMGGNPLE	N	*	*	188–200
P		GVFGLRN	M	*	*	181–188
K		NHLVEIPPNLPS <sup>#</sup>	S	*	*	152–163
K		LLQVVYL	H	*	*	305–311
N		GISLFNPNVPYW	E	*	*	338–349
P		YWEVQPA	T	*	*	348–354
COMP		N	WVVLNQGMEIVQTM <sup>#</sup>	N	*	*
	K	QVCTDIDECETGQHNCVPN <sup>#*3,7</sup>	S	*	*	174–192

B

Protein	Amino Acid Before	Peptide Sequence	Amino Acid After	Significant	Significant	Position
				MASCOT Identified Peptides; $p < 0.05$	PEAKS 10lgP	
COMP	R	AFQTVVLDPEGDAQIDPN <sup>#</sup>	W	*	*	531–548
	Q	CAPGSCFPVAC <sup>#*4,7</sup>	T	*	*	89–100
	R	WFLQHRPQVGY <sup>#</sup>	I	*	*	682–692
	R	NALWHTGDTASQ	V	*	*	650–661
	K	QVCTDIDECETGQHN	C	*	*	174–188
	K	QMEQTYWQANP	F	*	*	614–624
	R	AFQTVVLDPEGDAQIDPNWVVLNQGME	I	*	*	531–557
	R	NALWHTGDTAS	Q	*	*	650–659
	R	AFQTVVLDPEGDAQIDPNW	V	*	*	531–549
	R	AFQTVVLDPEGDAQIDPNWVVLN	Q	*	*	531–553
	N	WVVLNQGME	I	*	*	549–557
	R	AFQTVVLDPEGDAQIDPNWVVLNQGMEIVQTMN	S	*	*	531–563

TABLE 3  
Continued

B

Protein	Amino Acid Before	Peptide Sequence	Amino Acid After	Significant MASCOT Identified Peptides; $p < 0.05$	Significant PEAKS 10lgP	Position
Decorin	<b>N</b>	<b>SVCINTQ</b>	<b>C</b>			193–199
	<b>C</b>	<b>FSQENIIWAN</b>	<b>L</b>			725–734
	<b>I</b>	<b>DPNWWVLNQGMEIVQTMNSDPGLAVGYTA<sup>c</sup></b>	<b>F</b>			546–574
	<b>N</b>	<b>WVVLNQGMEIVQTMN</b>	<b>S</b>			549–563
	<b>R</b>	<b>IHEVLDLEPLGPVCPF<sup>#</sup></b>	<b>R</b>	*	*	42–57
Fibromodulin	<b>K</b>	<b>AVFNGLNQ</b>	<b>M</b>		*	169–176
	<b>K</b>	<b>ISPGAFPLV</b>	<b>K</b>		*	121–130
	<b>R</b>	<b>DCPQECDCPPNFPTAMY<sup>#</sup></b>	<b>C</b>	*	*	75–91
	<b>K</b>	<b>IPPVNTNLEN<sup>#</sup></b>	<b>L</b>		*	308–317
	<b>R</b>	<b>LSHNSLTNN<sup>#</sup></b>	<b>G</b>		*	275–283
	<b>R</b>	<b>INEFSISSFCTVVDVMN</b>	<b>F</b>	*	*	325–341
	<b>R</b>	<b>LSHNSLTNNGLAS</b>	<b>N</b>	*	*	275–287
	<b>R</b>	<b>KVPDGLPSALEQL</b>	<b>Y</b>	*	*	237–249
	<b>R</b>	<b>SLILLDLSYNH</b>	<b>L</b>	*	*	224–234
	<b>R</b>	<b>KVPDGLPSALE</b>	<b>Q</b>		*	237–247
	<b>R</b>	<b>SAMPADAPLC</b>	<b>L</b>		*	358–367
	<b>R</b>	<b>INEFSISSFC</b>	<b>T</b>		*	325–334
	<b>K</b>	<b>YVYFQNNQIS</b>	<b>S</b>			109–118
	<b>A</b>	<b>LEQLYLEHNNVY</b>	<b>S</b>			246–257
	<b>K</b>	<b>YVYFQNNQISS</b>	<b>I</b>			109–119
<b>R</b>	<b>KVPDGLPSA</b>	<b>L</b>		*	237–245	
<b>R</b>	<b>LSHNSLTNNG</b>	<b>L</b>		*	257–284	
<b>R</b>	<b>KVPDGLPS</b>	<b>A</b>			237–244	
<b>R</b>	<b>KVPDGLPSALEQLYLEHNNVYSVPDS</b>	<b>Y</b>			237–262	
<b>R</b>	<b>SAMPADAPLCL</b>	<b>R</b>			358–368	

In the peptide sequence column, \* denotes neopeptides produced at known cleavages sites to specific proteases at the N or C terminal end and # denotes neopeptide that have been identified following protease digestion with both ADAMTS-4 and MMP-3. A significant PEAKS<sup>®</sup> derived 10lg  $p$  value is  $>20$  that equates to  $p$  value  $<0.05$ . Potential cleavage sites are marked in bold. Position refers to the amino acid of the peptide sequence within the protein.<sup>\*1Ref.<sup>8</sup>; \*2Ref.<sup>15</sup>; \*3Ref.<sup>45</sup>; \*4Ref.<sup>7</sup>; \*5Ref.<sup>6</sup>; \*6(Holden P, personal communication); \*7Ref.<sup>49</sup>; \*8Ref.<sup>42</sup>; \*9Ref.<sup>44</sup>.</sup>

instrument. For instance, cleavage at PTPFKEEE<sup>1745</sup><sub>1746</sub>GLGSVELSG would produce a small peptide, K.EEE that is below the mass range identifiable and GLGSVELSGLPSGDADLSGTSGR whose mass would be too great for detection. Furthermore, neopeptides corresponding to other previously identified cleavage sites were evident in our data; six for MMP-3 treatment,<sup>7,8,42,44,45</sup> eight following ADAMTS-4 treatment<sup>6–8,15,45</sup> (Holden P, personal communication), and 15 in equine II-1 $\beta$  model<sup>6,7,44,45</sup> (Holden P, personal communication).

We studied the sequence context of the neopeptides within the corresponding protein. Interestingly, apart from decorin, where neopeptides identified were equally distributed throughout the protein, neopeptides generated from COMP, biglycan, and fibromodulin were primarily from the C-terminal region. This has previously been described for COMP.<sup>5</sup> This dominance of neopeptides with cleavage identified at the C terminus could be due to the structure and dynamics of the substrate protein responsible for the targeting in

the enzyme molecular recognition system.<sup>46</sup> Thus, there are likely sites of limited proteolysis that may account for our findings.

One of the limitations of this study is the lack of validation of candidate neopeptides. This would require the development of monoclonal epitope antibodies that was beyond the scope of this study. We were only able to validate the aggrecan neopeptide DIPEN. In addition, for some neopeptides, PEAKS<sup>®</sup> analysis was significant; whereas in Mascot, it was not (and visa-versa). We identified neopeptides with significant scores. This gives a probability that the sequence established by Mascot or PEAKS<sup>®</sup> was correct and the peptide mass and subsequent fragmentation pattern/de-novo sequencing were not due to another peptide. However, this is only a probability and some neopeptides may be either missed or falsely identified. An example of this is the identification of the aggrecan neopeptide<sup>1838</sup>EAGEGPSGILE. The expected aggrecanase-mediated cleavage is at<sup>1839</sup>AGEGPSGILE. This neopeptide has one potential chondroitin sulphate

**Table 4.** Neopeptides Identified Following Treatment of Cartilage Explants With IL-1 $\beta$

A		Amino Acid Before		Peptide Sequence	Amino Acid After	ADAMTS-4 Digest	MMP-3 Digest	Significant MASCOT	Significant PEAKS	Position
Aggrecan	<b>Q</b>	<b>E</b>	<b>A</b>	<b>EAGEGSPGILE</b>	<b>L</b>	✓				1838–1848
	<b>V</b>	<b>P</b>	<b>I</b>	<b>PIETELPSPGEPSPG</b>	<b>V</b>					793–806
Biglycan	<b>R</b>	<b>L</b>	<b>A</b>	<b>LAIQFGNY<sup>5</sup></b>	<b>K</b>			*	*	363–370
	<b>P</b>	<b>Y</b>	<b>W</b>	<b>YWEVQPATFRCVTDRLAIQ<sup>9</sup></b>	<b>F</b>					348–366
COMP	<b>R</b>	<b>D</b>	<b>V</b>	<b>DVDHDFVGDACDSQDKDGD</b>	<b>G</b>			*	*	418–437
	<b>R</b>	<b>D</b>	<b>V</b>	<b>DVDHDFVGDACDSQD</b>	<b>K</b>			*	*	418–433
	<b>R</b>	<b>C</b>	<b>E</b>	<b>CEACPPGYSGPTHEGVGM<sup>3</sup></b>	<b>A</b>			*	*	149–166
	<b>R</b>	<b>C</b>	<b>E</b>	<b>CEACPPGYSGPTHEGVGM<sup>6</sup>AF<sup>6</sup></b>	<b>A</b>	✓		*	*	149–168
	<b>R</b>	<b>F</b>	<b>C</b>	<b>FCPDGTFSPCHEKAD</b>	<b>C</b>			*	*	226–240
	<b>K</b>	<b>N</b>	<b>T</b>	<b>NTVMECDACGMQ<sup>3</sup></b>	<b>A</b>	✓		*	*	63–75
	<b>R</b>	<b>A</b>	<b>F</b>	<b>AFQTVVLDPEGDAQIDPNWVVLNQGMEIVQTMN</b>	<b>S</b>		✓		*	531–563
	<b>K</b>	<b>Q</b>	<b>M</b>	<b>QMEQTYWQA</b>	<b>N</b>			*	*	614–622
	<b>R</b>	<b>L</b>	<b>V</b>	<b>LVPNPGQEDADRDGVGD</b>	<b>V</b>			*	*	484–500
	<b>R</b>	<b>L</b>	<b>G</b>	<b>LVVFCFSQENIIWAN</b>	<b>L</b>		✓		*	720–734
A	<b>A</b>	<b>Q</b>	<b>C</b>	<b>QCAPGSCFPGVAC</b>	<b>T</b>		✓		*	88–100
	<b>R</b>	<b>N</b>	<b>A</b>	<b>NALWHITGD</b>	<b>T</b>				*	650–657
	<b>F</b>	<b>C</b>	<b>F</b>	<b>CFSQENIIW<sup>4</sup></b>	<b>A</b>				*	724–732
	<b>R</b>	<b>S</b>	<b>C</b>	<b>SCVCAVGWAGNGL</b>	<b>L</b>				*	250–262
	<b>R</b>	<b>N</b>	<b>A</b>	<b>NAVDCPRVPSDQKDSQD</b>	<b>D</b>				*	380–398
	<b>K</b>	<b>Q</b>	<b>V</b>	<b>QVCTDIDECEETGQHN</b>	<b>C</b>		✓	*	*	174–188
	<b>R</b>	<b>S</b>	<b>C</b>	<b>SCVCAVGWAGNGLC</b>	<b>G</b>			*	*	250–264
	<b>L</b>	<b>A</b>	<b>Q</b>	<b>AQCAPGSCFPGVACT<sup>5,3</sup></b>	<b>T</b>		✓	*	*	87–102
	<b>R</b>	<b>K</b>	<b>D</b>	<b>KDNCVTVPSNGQEDADRDGIGD<sup>3</sup></b>	<b>A</b>			*	*	287–308
	<b>R</b>	<b>V</b>	<b>S</b>	<b>VSVRPLAQCAPGSCFPGVA</b>	<b>C</b>			*	*	81–99
	<b>R</b>	<b>V</b>	<b>S</b>	<b>VSVRPLAQCAPGS</b>	<b>C</b>			*	*	81–93
	<b>Q</b>	<b>C</b>	<b>A</b>	<b>CAPGSCFPGVACTQTA<sup>4</sup></b>	<b>S</b>			*	*	89–104
<b>R</b>	<b>L</b>	<b>G</b>	<b>LVVFCFSQEN</b>	<b>I</b>			*	*	720–729	
<b>R</b>	<b>V</b>	<b>S</b>	<b>VSVRPLAQ<sup>4,5</sup></b>	<b>C</b>			*	*	81–88	
<b>R</b>	<b>V</b>	<b>S</b>	<b>VSVRPLAQ</b>	<b>A</b>			*	*	81–89	
B		Amino Acid Before	Peptide Sequence	Amino Acid After	ADAMTS-4 Digest	MMP-3 Digest	Significant MASCOT	Significant PEAKS	Position	
Protein	<b>R</b>	<b>A</b>	<b>AFQTVVLDPEGDAQIDPNWVVLN<sup>Q</sup></b>	<b>G</b>			*	*	531–554	
	<b>Q</b>	<b>C</b>	<b>CAPGSCFPGVAC<sup>4,7</sup></b>	<b>T</b>			*	*	89–100	
	<b>Q</b>	<b>C</b>	<b>CAPGSCFPGVACT<sup>4</sup>Q</b>	<b>T</b>			*	*	89–102	

TABLE 4  
Continued

Protein	Amino Acid Before	Peptide Sequence	Amino Acid After	ADAMTS-4 Digest	MMP-3 Digest	Significant MASCOT	Significant PEAKS	Position
COMP	K	QVCTDIDECETGQHNCVVPN	S		✓	*	*	174–192
	R	VPNSDQKDSGD	G			*	*	388–399
	Q	CAPGSCFFGVACT* <sup>4</sup>	Q			*	*	89–101
	R	VSVRPLAQCAPGSCFFGVACTQT	A			*	*	81–103
	R	LGVFCFSQE	N			*	*	720–728
	R	KDNCVIVPNSGQED	A			*	*	287–300
	R	SCVCAVGW	A			*	*	250–257
	R	SCVCAVGWAGNGLL	C			*	*	250–263
	R	NALWHTGDTASQ	V		✓	*	*	650–661
	R	SCVCAVGWAGN	G			*	*	250–260
	C	AVGWAGNGLLC	G			*	*	254–264
	R	NAVDCNCRVPSND	Q			*	*	380–392
	R	FYEGPELVADSNVLLD	T			*	*	697–712
	Q	CAPGSCFFGVA* <sup>4</sup>	C			*	*	89–99
	R	LVPNPGQEDADRD	G			*	*	484–496
	M	ECDACGMQPAR	T			*	*	67–77
	R	DVDHDFVGDACD	S			*	*	418–429
	R	NTDGDKWGD	A			*	*	336–344
R	VSVRPLAQCAPG	S			*	*	81–92	
R	VSVRPLAQCAPGSCFFGVAC	T			*	*	81–100	
A	VGWAGNGLLCGR	D			*	*	255–266	
K	KASYSGVSLF	S			*	*	322–331	
Q	MIVVELGTNPLKSSG* <sup>3</sup>	I		✓			177–191	
P	PGLPPSLTEL	H			*	*	218–227	
R	KVPDGLPSALEQL	Y			*	*	237–249	
R	INEFSISSFC* <sup>7</sup> TVVDVM	N			*	*	325–340	
R	KVPDGLPSALEQLYLEHNNVY	S		✓			237–257	

Table indicates neopeptides identified in at least two donor samples. Neopeptides also evident in the ADAMTS-4 and MMP-3 experiments are also indicated with a tick. \* indicates neopeptides produced at known cleavages sites to specific proteases. Peptides with a significant ( $p < 0.05$ ) match that the sequence was correct based on Mascot ion scores are indicated with a black asterisk. A significant PEAKS<sup>®</sup> derived 10lg  $p$  value is  $> 20$ ; equating to a  $p$ -value  $< 0.05$ . Position refers to the amino acid of the peptide sequence within the protein. \*<sup>4</sup>Ref. <sup>7</sup>; \*<sup>5</sup>Ref. <sup>6</sup>; \*<sup>6</sup>(Holden P, personal communication); \*<sup>7</sup>Ref. <sup>49</sup>; \*<sup>9</sup>Ref. <sup>44</sup>

**Table 5.** Summarised Neopeptides Data

Protein	Amino Acid Before Peptide Residue	Peptide Sequence	Amino Acid Residue After	ADAMTS-4 Digest	MMP-3 Digest	IL-1 Treatment	Position
Aggrecan	V	EDISGFPSGGE	V	✓	✓		989-998
	R	LATTGQLYLAW	Q	✓	✓		280-290
	R	LATTGQLY	L	✓	✓		280-287
	R	VSLPNYPAIPTDATLELN	L	✓	✓		105-122
	R	VSLPNYPAIPTDATLE	L	✓	✓		105-119
	R	WSDGHSLQFEN	W	✓	✓		2275-2285
	R	YDAICYTGEDFVDIPEN <sup>*8</sup>	F	✓	✓		344-380
	Q	EAGEGPSGILE	L	✓		✓	1838-1848
	N	CIEMGGNPLENSGFQPGAFDGLK <sup>*2,3</sup>	L	✓			191-213
	N	GISLFNNVPY <sup>*3</sup>	W	✓	✓		338-348
Biglycan	K	LLQVVYLH	T	✓	✓		305-312
	K	NHLVEPPNLPSS <sup>*1</sup>	L	✓	✓		152-164
	K	NHLVEPPNLPSS	S	✓	✓		152-163
	K	VGVNDFCPVGFV	K	✓	✓		319-328
	Y	NGISLFNNVP <sup>*9</sup>	Y	✓	✓		337-347
	R	LAIQFGNY <sup>*5</sup>	K			✓	363-370
	P	YWEVQPATFRVCVTDRLAIQ <sup>*9</sup>	F			✓	348-366
	L	AQCAPGSCFPGVACTQ <sup>*5,3</sup>	T	✓	✓	✓	87-102
	R	AFQTVVLDPEGDAQIDPNWVVLNQGMEIVQTMN	S	✓	✓	✓	531-563
	R	AFQTVVLDPEGDAQIDPN	W	✓	✓		531-548
COMP	Q	CAPGSCFPGVA <sup>*4</sup>	C			✓	89-99
	Q	CAPGSCFPGVAC <sup>*4,7</sup>	T			✓	89-100
	Q	CAPGSCFPGVACT <sup>*4</sup>	Q			✓	89-101
	Q	CAPGSCFPGVACTQ <sup>*4</sup>	T			✓	89-102
	Q	CAPGSCFPGVACTQTA <sup>*4</sup>	S			✓	89-104
	R	CEACPPGYSGPTHEGVGMAF <sup>*6</sup>	A	✓		✓	149-168
	F	CFSQENIIW <sup>*4</sup>	A			✓	724-732
	R	KDNCVIVPNSGGEDADRDGIGD <sup>*8</sup>	A			✓	287-308
	R	LGVFCFSQENIIWAN	L		✓	✓	720-734
	R	NALWHTGDTASQ	V		✓	✓	650-661
Decorin	K	NTVMECDACGMQP <sup>*5</sup>	A	✓	✓		63-75
	K	QVCTDIDECETGQHNCVPN <sup>*3,7</sup>	S	✓	✓		174-192
	K	QVCTDIDECETGQHN	C	✓	✓		174-188
	R	VSVRPLAQ <sup>*4,5</sup>	C			✓	81-88
	R	WFLQHRPQVGY	I	✓	✓		682-692
	R	IHEVLDLEPLGPVCPFF	R	✓	✓		42-57

TABLE 5  
Continued

Protein	Amino Acid Before Peptide Residue	Peptide Sequence	Amino Acid Residue After	ADAMTS-4 Digest	MMP-3 Digest	IL-1 Treatment	Position
Fibromodulin	Q	MIVVELGTNPLKSSG* <sup>3</sup>	I	✓		✓	177-191
	R	DCPQECDCPPNFTAMY	C	✓	✓		75-91
	R	ELHLDHN* <sup>1</sup>	Q	✓			180-186
	K	IPPVNTNLEN	L	✓	✓		308-317
	R	KVPDGLPSALEQLYLEHNNVY	S	✓	✓	✓	237-257
	R	KVPDGLPSALEQL	Y	✓	✓	✓	237-249
	R	LSHNSLTNN	G	✓	✓		275-283

Position refers to the amino acid of the peptide sequence within the protein. \*Neopeptides produced at known cleavages sites to specific proteases.

attachment site at SG and the workflow used would not fully remove this glycosaminoglycan side chain. The presence of a CS would significantly alter the mass of the neopeptide and thus preclude identification. Although it is likely that there is a CS at this site, we cannot be totally confident about this without undertaking further work. If a CS was present here, it would change the mass of the peptide and not be identified by our analysis.

Although protease inhibitors were present throughout our workflow, it is possible that some of the peptides identified were cleaved by endogenous proteases. This is especially likely where peptides are seen with successive amino acids removed, such as the exoproteolytic fraying of <sup>152</sup>NHLVEIPPNLPS to <sup>152</sup>NHLVEIPPNLPS produced following MMP3 treatment in biglycan. Most peptides produced by endogenous proteases were eliminated from the data analysis as peptides identified in both control and protease-treated conditions were excluded.

Other potential anomalies in the data are regarding miscleaves due to the digestion enzyme trypsin. Trypsin cleaves the carboxy-terminal peptide bond of both Arg and Lys. However, cleavage by trypsin is not always reproducible or predictable. For example, one of the most usual mistakes is the omission of a cleavable residue (miscleavage). For trypsin, such miscleaved positions can be predicted when an Arg/Lys is followed by Pro, when successive Lys/Arg or positive charges are close to each other, and if several Asp/Glu are close to the positively charged residue. For example, the decorin neopeptide following MMP3 and ADAMTS4 treatment <sup>42</sup>IHEVL DLEPLGPVCPF.R. This semitryptic neopeptide (at the C terminus) is cleaved between Phe and Arg. This could be due to a miscleave due to the presence of upstream Pro.

There is a potential for some unique neopeptides identified to act as markers of specific proteases, cartilage degradation, and to measure ECM breakdown in other tissues.<sup>6,7,44</sup> Until recently, the progression of OA has been measured almost exclusively by radiography to assess space narrowing, or by magnetic resonance imaging of the joint. There is a need to progress to translatable biomarkers with the ability to monitor protease activity and OA disease. The peptide fragments demonstrated in this study could serve as early indicators of cartilage turnover similar to others studied in COMP<sup>47</sup> and aggrecan<sup>48</sup> by providing candidates for assays in animal studies and clinical settings, not only in cartilage but wherever these proteins are expressed, such as tendon, ligament, muscle, and liver. In our recent synovial fluid study, we identified some of the neopeptides identified in this study in synovial fluid from OA but not normal horses,<sup>45</sup> providing further evidence that some of the neopeptides identified here could provide biomarkers. Indeed the neopeptides identified in this study and in early OA synovial fluid would be ideal candidates for future studies to delineate

degradative mechanism in-vivo. For example, the biglycan neopeptides <sup>348</sup>YWEVQPATFRCVTDRLAIQ and <sup>191</sup>CIEMGGNPLENSGFQPGAFDGLK and the COMP neopeptide <sup>174</sup>QVCTDIDECETGQHN are well supported in the literature and may be the candidates for raising monoclonal neo-epitope antibodies. Furthermore, a number of biglycan, aggrecan, COMP, decorin, and fibromodulin neopeptides following  $\text{IL-1}\beta$  treatment of cartilage explants, and/or ADAMTS4 or MMP3 treatment of cartilage crude proteoglycan extract were identified for the first time. Many of the sequences are highly conserved between equine, human, and bovine. These are potential candidates for generating antibodies that could assist in elucidating the role of proteases in not only cartilage matrix protein turnover but also in other tissues, such as tendon. Furthermore, with ever increasing elegant mass spectrometry techniques becoming accessible in the future, it may be possible to use defined sets of the neopeptides themselves as markers of cartilage degradation and OA. Our neopeptides could lead to novel methods for OA biomarker pattern identification, validation, and more importantly OA patient stratification data to facilitate appropriate OA diagnosis and treatment.

## CONCLUSIONS

Improved knowledge of specific peptide fragments and the proteases generating these fragments will aid in the identification of markers of ECM degradation such as in OA.

## AUTHORS' CONTRIBUTIONS

M.J.P. conceived the study, carried out all the laboratory work and analysis, and drafted the manuscript. P.D.C. and D.J.T. participated in the design and coordination of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGEMENTS

M.J.P. was supported by a Wellcome Trust Veterinary Integrated Research Fellowship (WT088557MA). We thank Timothy Koh Jia Wei for his assistance with producing the tables for this manuscript.

## REFERENCES

1. Poole AR, Kojima T, Yasuda T, et al. 2001. Composition and structure of articular cartilage: a template for tissue repair. *Clin Orthop Relat Res* (391 Suppl):S26–S33.
2. Martel-Pelletier J, McCollum R, Fujimoto N, et al. 1994. Excess of metalloproteases over tissue inhibitor of metalloprotease may contribute to cartilage degradation in osteoarthritis and rheumatoid arthritis. *Lab Invest* 70:807–815.
3. Poole AR, Rizkalla G, Ionescu M, et al. 1993. Osteoarthritis in the human knee: a dynamic process of cartilage matrix degradation, synthesis and reorganization. *Agents Actions Suppl* 39:3–13.
4. Kempson GE, Muir H, Pollard C, et al. 1973. The tensile properties of the cartilage of human femoral condyles related to the content of collagen and glycosaminoglycans. *Biochim Biophys Acta* 297:456–472.
5. Zhen EY, Brittain IJ, Laska DA, et al. 2008. Characterization of metalloprotease cleavage products of human articular cartilage. *Arthritis Rheum* 58:2420–2431.
6. Peffers MJ, Thorpe CT, Collins JA, et al. 2014. Proteomic analysis reveals age-related changes in tendon matrix composition, with age- and injury-specific matrix fragmentation. *J Biol Chem* 289:25867–25878.
7. Dakin SG, Smith RK, Heinegard D, et al. 2014. Proteomic analysis of tendon extracellular matrix reveals disease stage-specific fragmentation and differential cleavage of COMP (cartilage oligomeric matrix protein). *J Biol Chem* 289:4919–4927.
8. Peffers MJ, Cillero-Pastor B, Eijkel GB, et al. 2014. Matrix assisted laser desorption ionization mass spectrometry imaging identifies markers of ageing and osteoarthritic cartilage. *Arthritis Res Ther* 16:R110.
9. Shohani B, Orazizadeh M, Hashemitabar M, et al. 2010. Degradation of extracellular matrix molecules in interleukin-1 $\alpha$  treated bovine nasal cartilage. *Iran Biomed J* 14:158–163.
10. Heinegard D. 2009. Proteoglycans and more-from molecules to biology. *Int J Exp Pathol* 90:575–586.
11. Nagase H, Kashiwagi M. 2003. Aggrecanases and cartilage matrix degradation. *Arthritis Res Ther* 5:94–103.
12. Fosang AJ, Rogerson FM, East CJ, et al. 2008. ADAMTS-5: the story so far. *Eur Cell Mater* 15:11–26.
13. Song RH, Tortorella MD, Malfait AM, et al. 2007. Aggrecan degradation in human articular cartilage explants is mediated by both ADAMTS-4 and ADAMTS-5. *Arthritis Rheum* 56:575–585.
14. Monfort J, Tardif G, Reboul P, et al. 2006. Degradation of small leucine-rich repeat proteoglycans by matrix metalloprotease-13: identification of a new biglycan cleavage site. *Arthritis Res Ther* 8:R26.
15. Melching LI, Fisher WD, Lee ER, et al. 2006. The cleavage of biglycan by aggrecanases. *Osteoarthritis Cartilage* 14:1147–1154.
16. Imai K, Hiramatsu A, Fukushima D, et al. 1997. Degradation of decorin by matrix metalloproteinases: identification of the cleavage sites, kinetic analyses and transforming growth factor- $\beta$ 1 release. *Biochem J* 322:809–814.
17. Heathfield TF, Onnerfjord P, Dahlberg L, et al. 2004. Cleavage of fibromodulin in cartilage explants involves removal of the N-terminal tyrosine sulfate-rich region by proteolysis at a site that is sensitive to matrix metalloproteinase-13. *J Biol Chem* 279:6286–6295.
18. Dickinson SC, Vankemmelbeke MN, Buttle DJ, et al. 2003. Cleavage of cartilage oligomeric matrix protein (thrombospondin-5) by matrix metalloproteinases and a disintegrin and metalloproteinase with thrombospondin motifs. *Matrix Biol* 22:267–278.
19. Rousseau JC, Delmas PD. 2007. Biological markers in osteoarthritis. *Nat Clin Pract Rheumatol* 3:346–356.
20. Kawcak CE, Frisbie DD, Werpy NM, et al. 2008. Effects of exercise vs experimental osteoarthritis on imaging outcomes. *Osteoarthritis Cartilage* 16:1519–1525.
21. Peffers MJ, Beynon RJ, Clegg PD. 2013. Absolute quantification of selected proteins in the human osteoarthritic secretome. *Int J Mol Sci* 14:20658–20681.
22. Heinegard D, Axelsson I. 1977. Distribution of keratan sulfate in cartilage proteoglycans. *J Biol Chem* 252:1971–1979.
23. Collins JA, Moots RJ, Winstanley R, et al. 2013. Oxygen and pH-sensitivity of human osteoarthritic chondrocytes in 3-D alginate bead culture system. *Osteoarthritis Cartilage* 21:1790–1798.

24. Aksoy N, Thornton DJ, Corfield A, et al. 1999. A study of the intracellular and secreted forms of the MUC2 mucin from the PC/AA intestinal cell line. *Glycobiology* 9:739–746.
25. Yan J, Liu L, Li X, et al. 2006. [A potential use of collagen-hyaluronan-chondroitin sulfate tri-copolymer scaffold for cartilage tissue engineering]. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi* 20:130–133.
26. Kokebie R, Aggarwal R, Lidder S, et al. 2011. The role of synovial fluid markers of catabolism and anabolism in osteoarthritis, rheumatoid arthritis and asymptomatic organ donors. *Arthritis Res Ther* 13:R50.
27. Chevalier X, Giraudeau B, Conrozier T, et al. 2005. Safety study of intraarticular injection of interleukin 1 receptor antagonist in patients with painful knee osteoarthritis: a multicenter study. *J Rheumatol* 32:1317–1323.
28. Cohen SB, Proudman S, Kivitz AJ, et al. 2011. A randomized, double-blind study of AMG 108 (a fully human monoclonal antibody to IL-1R1) in patients with osteoarthritis of the knee. *Arthritis Res Ther* 13:R125.
29. Bougault C, Gosset M, Houard X, et al. 2012. Stress-induced cartilage degradation does not depend on the NLRP3 inflammasome in human osteoarthritis and mouse models. *Arthritis Rheum* 64:3972–3981.
30. Clements KM, Price JS, Chambers MG, et al. 2003. Gene deletion of either interleukin-1beta, interleukin-1beta-converting enzyme, inducible nitric oxide synthase, or stromelysin 1 accelerates the development of knee osteoarthritis in mice after surgical transection of the medial collateral ligament and partial medial meniscectomy. *Arthritis Rheum* 48:3452–3463.
31. Jotanovic Z, Mihelic R, Sestan B, et al. 2012. Role of interleukin-1 inhibitors in osteoarthritis: an evidence-based review. *Drugs Aging* 29:343–358.
32. Fosang AJ, Tyler JA, Hardingham TE. 1991. Effect of interleukin-1 and insulin like growth factor-1 on the release of proteoglycan components and hyaluronan from pig articular cartilage in explant culture. *Matrix* 11:17–24.
33. Kobayashi M, Squires GR, Mousa A, et al. 2005. Role of interleukin-1 and tumor necrosis factor alpha in matrix degradation of human osteoarthritic cartilage. *Arthritis Rheum* 52:128–135.
34. Lark MW, Bayne EK, Flanagan J, et al. 1997. Aggrecan degradation in human cartilage. Evidence for both matrix metalloproteinase and aggrecanase activity in normal, osteoarthritic, and rheumatoid joints. *J Clin Invest* 100:93–106.
35. Goodrich LR, Nixon AJ. 2006. Medical treatment of osteoarthritis in the horse—a review. *Vet J* 171:51–69.
36. Tortorella MD, Malfait AM. 2008. Will the real aggrecanase(s) step up: evaluating the criteria that define aggrecanase activity in osteoarthritis. *Curr Pharm Biotechnol* 9:16–23.
37. Shinmei M, Masuda K, Kikuchi T, et al. 1991. Production of cytokines by chondrocytes and its role in proteoglycan degradation. *J Rheumatol Suppl* 27:89–91.
38. Gao G, Plaas A, Thompson VP, et al. 2004. ADAMTS4 (aggrecanase-1) activation on the cell surface involves C-terminal cleavage by glycosylphosphatidyl inositol-anchored membrane type 4-matrix metalloproteinase and binding of the activated proteinase to chondroitin sulfate and heparan sulfate on syndecan-1. *J Biol Chem* 279:10042–10051.
39. Tseng S, Reddi AH, Di Cesare PE. 2009. Cartilage oligomeric matrix protein (COMP): a biomarker of arthritis. *Biomark Insights* 4:33–44.
40. Zhang J, Xin L, Shan B, et al. 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* 11:M111.010587.
41. Siepen JA, Keevil EJ, Knight D, et al. 2007. Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J Proteome Res* 6:399–408.
42. Flannery CR, Lark MW, Sandy JD. 1992. Identification of a stromelysin cleavage site within the interglobular domain of human aggrecan. Evidence for proteolysis at this site in vivo in human articular cartilage. *J Biol Chem* 267:1008–1014.
43. Roughley PJ, Geng Y, Mort JS. 2014. The non-aggregated aggrecan in the human intervertebral disc can arise by a non-proteolytic mechanism. *Eur Cell Mater* 28:136.
44. Genovese F, Barasuc N, Larsen L, et al. 2013. Biglycan fragmentation in pathologies associated with extracellular matrix remodeling by matrix metalloproteinases. *Fibrogenesis Tissue Repair* 6:9.
45. Peffers MJ, Mcdermott, BM, Clegg, PD, et al. 2015. Comprehensive protein profiling of synovial fluid in osteoarthritis following protein equalization. *Osteoarthritis Cartilage* 23:1204–1213.
46. Hubbard SJ, Beynon RJ, Thornton JM. 1998. Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures. *Protein Eng* 11:349–359.
47. Hunter DJ, Lavalley M, Li J, et al. 2007. Urinary pentosidine does not predict cartilage loss among subjects with symptomatic knee OA: the BOKS Study. *Osteoarthritis Cartilage* 15:93–97.
48. Dufield DR, Nemirovskiy OV, Jennings MG, et al. 2010. An immunoaffinity liquid chromatography-tandem mass spectrometry assay for detection of endogenous aggrecan fragments in biological fluids: use as a biomarker for aggrecanase activity and cartilage degradation. *Anal Biochem* 406:113–123.
49. Ahrman E, Lorenzo P, Holmgren K, et al. 2014. Novel cartilage oligomeric matrix protein (COMP) neopeptides identified in synovial fluids from patients with joint diseases using affinity chromatography and mass spectrometry. *J Biol Chem* 289:20908–20916.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.