Osteoarthritis and Cartilage 25 (2017) 1496–1504

Osteoarthritis and Cartilage

Cartilage oligomeric matrix protein forms protein complexes with synovial lubricin via non-covalent and covalent interactions

S.A. Flowers † [a], S. Kalamajski ‡ [a], L. Ali †, L.I. Björkman §, J.R. Raj †, A. Aspberg ‡, N.G. Karlsson † * [b], C. Jin † [b]

† *Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*
‡ *Department of Clinical Sciences Lund, Division of Rheumatology and Molecular Skeletal Biology, Lund University, Lund, Sweden*
§ *Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*

ARTICLE INFO

Article history:
Received 9 August 2016
Accepted 23 March 2017

Keywords:
Boundary lubrication
Cartilage degradation
Proteomics
O-linked glycoproteins
Lubricin
Cartilage oligomeric matrix protein

SUMMARY

Objective: Understanding the cartilage surface structure, lost in arthritic disease, is essential for developing strategies to effectively restore it. Given that adherence of the lubricating protein, lubricin, to the cartilage surface is critical for boundary lubrication, an interaction with cartilage oligomeric matrix protein (COMP) was investigated. COMP, an abundant cartilage protein, is known to be important for matrix formation.

Design: Synovial fluid (SF) from arthritic patients was used to detect possible COMP–lubricin complexes by immunological methods. Recombinant (RC) COMP and lubricin fragments were expressed to characterize this bonding and mass spectrometry employed to specifically identify the cysteines involved in inter-protein disulfide bonds.

Results: COMP–lubricin complexes were identified in the SF of arthritic patients by Western blot, co-immunoprecipitation and sandwich ELISA. RC fragment solid-phase binding assays showed that the *C*-terminal (amino acids (AA) 518-757) of COMP bound non-covalently to the *N*-terminal of lubricin (AA 105-202). Mass spectrometry determined that although cysteines throughout COMP were involved in binding with lubricin, the cysteines in lubricin were primarily focused to an *N*-terminal region (AA 64-86). The close proximity of the non-covalent and disulfide binding domains on lubricin suggest a two-step mechanism to strongly bind lubricin to COMP.

Conclusion: These data demonstrate that lubricin forms a complex network with COMP involving both non-covalent and covalent bonds. This complex between lubricin and the cartilage protein COMP can be identified in the SF of patients with arthritis conditions including osteoarthritis (OA) and rheumatoid arthritis (RA).

Introduction

Arthritic diseases including osteoarthritis (OA)[1] and rheumatoid arthritis (RA)[2] result in destruction of the joint surface leading to debilitating pain and restricted motion. Although RA treatment is increasing in efficacy[3], OA treatment is limited to symptom management, making understanding joint biomechanics essential for devising treatments to re-establish the lubricating joint surface.

The cartilage extracellular matrix (ECM) in diarthrodial joints has lubricating and shock-absorbing functions. Lubrication is created by two modes: *boundary mode*, created at the cartilage surface, and *mixed mode*, a combination of boundary and hydrodynamic mode lubrication[4,5]. Hydrodynamic mode is generated by major synovial fluid (SF) components including hyaluronic acid and lubricin, a heavily *O*-glycosylated, mucin-like glycoprotein[4,5]. Boundary lubrication at the mechanically vulnerable cartilage surface creates a smooth non-adherent low friction coating vital for cartilage surface structural integrity; studies show that lubricin is essential for these properties[4,6,7].

* Address correspondence and reprint requests to: N.G. Karlsson, Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Medicinaregatan 9A, 405 30 Gothenburg, Sweden. Fax: 46-31-7862150.

E-mail addresses: sarah.flowers@medkem.gu.se (S.A. Flowers), sebastian.kalamajski@imbim.uu.se (S. Kalamajski), l.z.ali@dundee.ac.uk (L. Ali), lena.i.bjorkman@vgregion.se (L.I. Björkman), jaison_505@yahoo.co.in (J.R. Raj), anders.aspberg@med.lu.se (A. Aspberg), niclas.karlsson@medkem.gu.se (N.G. Karlsson), chunsheng.jin@medkem.gu.se (C. Jin).

[a] Co-first author.
[b] Co-last author.

http://dx.doi.org/10.1016/j.joca.2017.03.016

Lubricin (PRG4), also known as superficial zone protein and proteoglycan 4, has an expected molecular weight of 150 kDa that is almost doubled by its glycosylation. Although also present in the blood, urine and on the ocular surface[8–10], in the joint, lubricin is synthesized by articular chondrocytes and synoviocytes, and enriched in the superficial zone of articular cartilage, the synovial membrane, tendons and the SF[11,12]. Although it seems that adherence of lubricin to the cartilage surface is necessary for boundary lubrication, the assembly remains unknown.

The SF of arthritic patients provides a unique opportunity to analyse this surface structure as it is lost from the cartilage into the SF during disease. Proteomic analysis of the acidic fraction of SF has identified lubricin and a range of other proteins, including cartilage oligomeric matrix protein (COMP)/thrombospondin 5 (TSP5)[13,14]. COMP is a homopentameric ECM glycoprotein, with five 100–110 kDa subunits[15]. COMP is mainly synthesized by chondrocytes, secreted into cartilage ECM where it is essential for matrix formation through its interactions with a range of molecules including aggrecan, fibronectin, matrilins and collagens[16–20]. COMP and its fragments are released into the SF, where they can interact with and regulate complement factors[21]. COMP in serum is considered a biomarker for early cartilage damage[22].

As COMP has complex forming capabilities and may co-purify with lubricin, we investigated a potential COMP–lubricin complex in arthritic SF. These analyses were able to identify a complex network of non-covalent and covalently bonded COMP and lubricin.

Methods

Patient sample information and study design

Eight independent SF samples from RA ($n = 5$ designated RA1, RA2, RA3, RA4, and RA5), OA ($n = 2$, OA1 and OA2) and Spondyloarthritis (SpA) ($n = 1$, SpA1) patients were collected during aspiration of knee joints at the Rheumatology Clinic, Sahlgrenska University Hospital (Gothenburg, Sweden) and Lund University Hospital (Lund, Sweden). All RA patients fulfilled the American College of Rheumatology 1987 revised criteria for RA[23]. All patients gave informed consent and the procedure approved by the Ethics Committees of Gothenburg and Lund Universities. Acidic fractions, containing lubricin and COMP, were purified from SF as previously described[13].

Discovery Western blot, proteomic, co-immunoprecipitation, size exclusion chromatography and ELISA experiments used human SF samples (*n* described throughout results and dependent on sample volume availability). The binding domain of each protein involved in bond formation was then determined by solid-phase binding assays using truncated recombinant (RC) COMP and lubricin, and MS disulfide bond analyses of SF ($n = 2$).

Immunoassays of synovial lubricin and COMP

Western blotting

Samples were separated using agarose-SDS-PAGE or 3–8% Tris/acetate gels. Transferred PVDF membranes were blocked (3% bovine serum albumin (BSA) in PBS), and probed with mouse anti-human lubricin (mAb 13, Pfizer Research), or rat anti-human COMP (mAb HC484D1, AbD Serotec). HRP conjugated secondary antibodies included rabbit anti-mouse immunoglobulins (DakoCytomation) and rabbit anti-rat IgG and IgM (Jackson ImmunoResearch Europe Ltd.).

Co-immunoprecipitation

SF acidic fractions (8 μg) and anti-COMP antibody or PBS were incubated 4°C overnight. Protein G agarose beads (100 μL, Thermo Scientific) were added for 2 h at RT, washed (3 × 10 min 0.75% Triton X-100 then PBS 4 × 10 min) and resuspended in non-reducing sample buffer for SDS-PAGE and Western blot analysis using anti-lubricin mAb13. ImageJ was used to quantitatively compare bands[26].

Size exclusion chromatograpy

HPLC with a BioSuite 450 column (8 μm HR SEC, 7.8 × 300 mm, Waters) was used to separate a portion of enriched SF glycoproteins, at a flow rate of 0.25 mL/min with a mobile phase of 2 M urea in PBS. The eluents (0.25 mL/fraction) were then analyzed by ELISA using mouse anti-human lubricin-specific mAb (mAb13, Pfizer Research) and rat anti-human COMP-specific mAb (mAb HC484D1, AbD Serotec).

Sandwich ELISA

Anti-COMP mAb (HC484D1, 10 μL/mL in 50 mM sodium carbonate buffer, pH 9.6, 16 h, 4°C) was the capture antibody and coated onto 96 well plate, blocked (1% BSA in PBS-T) and incubated with diluted SF. Anti-lubricin (mAb13) was added and detected with HRP conjugated rabbit anti-mouse Ig at 0.5 μg/mL and 3,3′,5,5′-tetramethylbenzidine before absorbance was measured (450 nm). Assay performed in triplicate.

Solid-phase binding assays

Assays were conducted with conditions as for sandwich ELISAs. RC protein (5 μg/mL) coated plates were incubated with potential interacting proteins (1 μg/mL) and binding detected with mouse anti-His mAb (18184, AbCam) and alkaline phosphatase-conjugated goat anti-mouse and visualized using para-nitrophenylphosphate. Assays were performed in triplicate and repeated twice, statistics described below.

Production of RC lubricin and COMP fragments

RC human lubricin with a FLAG-tag and a truncated mucin-like domain (without amino acids (AA) 403-870) was produced in 293F cells using p3xFLAG-CMV-8 vector (Sigma–Aldrich) and purified on anti-FLAG beads. The same method was used to produce four lubricin fragments named L25-221 (all molecular weights are calculated, MW 21.9 kDa), L220-402 (MW 19.2 kDa), L871-1078 (MW 22.6 kDa), L1079-1404 (MW 37.2 kDa, Fig. 2(A)). GST-tagged fragments of the *N*-terminal domain divided by exon boundaries (exons 2–5), named L25-72 (MW 5.3 kDa), L67-109 (MW 5.0 kDa), L105-160 (MW 6.0 kDa), L155-202 (MW 5.2 kDa), were produced using pGEX-5X-3 vector (GE Healthcare) and Rosetta 2 *E. coli* (Novagen), purified in native conditions using glutathione beads (Pierce). The His-tagged fragments of COMP [Fig. 2(A)] named mCOMP (AA 73-757, MW 79.3 kDa), TII1-TIII8 (AA 73-517, MW 51.7 kDa), TIII1-CG (AA 268-757, MW 59.0 kDa), TIII1-8 (AA 268-517, MW 31.5 kDa) and CG (AA 518-757, MW 31.9 kDa), based on the AA sequence of the human COMP reference sequence (NM_000095.2), were produced in human 293-EBNA cells and purified as previously described[21]. The identity of the proteins was confirmed by mass spectrometry, and purity by SDS-PAGE (Fig. S3).

Mass spectrometry methods

Protein identification from gel bands

Coomassie stained gel bands were excised, destained, dried and digested with 300 ng sequencing-grade lysine-C (50 mM ammonium bicarbonate, 4 h, 37°C, Promega) then 500 ng trypsin (overnight, 37°C, Promega). Samples were dried and resuspended in 50 mM ammonium bicarbonate with 0.2% formic acid for LC-MS/MS analysis. The MassMatrix conversion tool was used for file

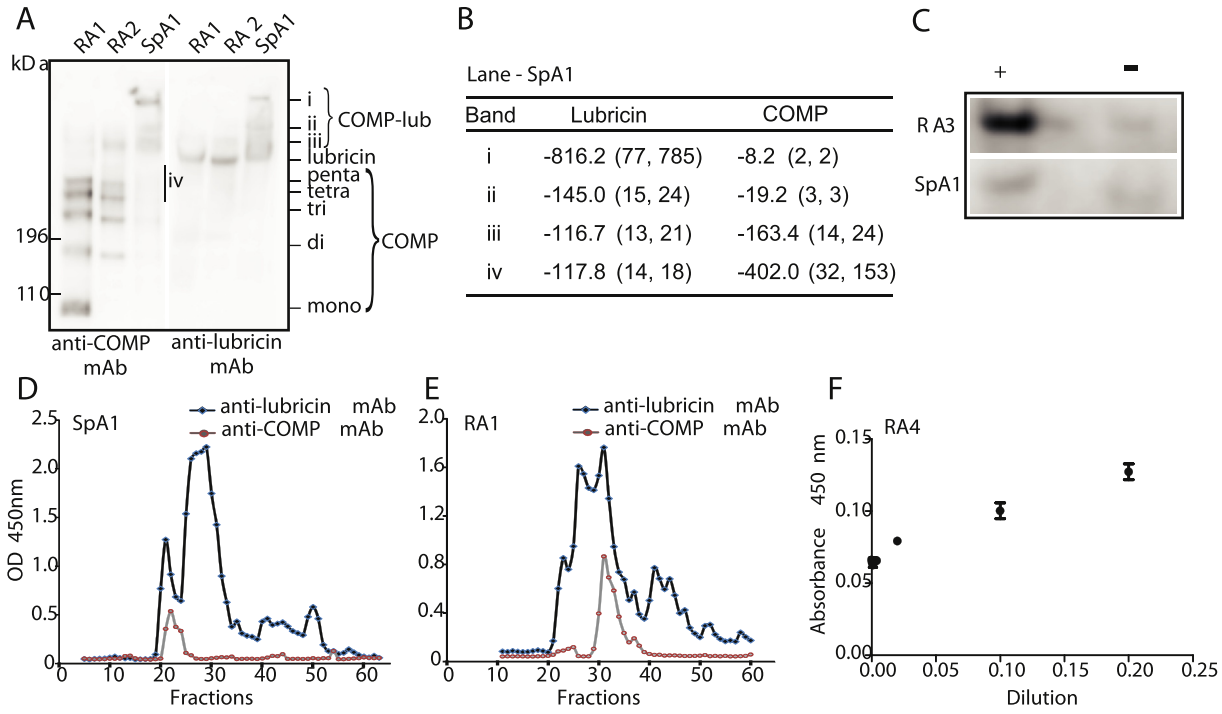

Fig. 1. Complexes between COMP and lubricin in the SF of arthritis patients. (A) Western blots comparing enriched glycoproteins from arthritis patients' SF. One SpA (SpA1) and two RA (RA1 and RA2) patient samples separated by AgSDS-PAGE, non-reducing conditions. After blotting to PVDF, lubricin was detected by mAb13; COMP detected by mAb HC484D1. (B) MS identification of proteins from the four HMW bands from SpA1. The confidence of identification (ID) from X! Tandem shown for lubricin and COMP for each band as log(e), the base-10 log of the expectation value, that is, the probability the ID is a random assignment. In brackets, the numbers of unique and total peptides used in the ID. (C) Co-IP of synovial COMP–lubricin complex from the acidic fraction of SF from patients RA3 and SpA1 (+). mAb HC484D1) was used, with protein G beads, to pull out the COMP–lubricin complex. After SDS-PAGE, a Western blot was used with mAb13 to identify the complex. A no antibody control (−) shows the complex enriched above non-specific binding. (D and E) Size exclusion chromatography of enriched glycoproteins from SF (D, patient SpA1; E, patient RA1) showing COMP and lubricin were co-purified. Chromatographic fractions were analyzed by ELISA with mAb13 and mAb HC484D1. (F) Sandwich ELISA of serial dilutions of acidic SF glycoproteins from patient RA4. mAb HC484D1 was used as the capture antibody and the level of tethered lubricin was detected by mAb13 and HRP conjugated rabbit anti-mouse antibody. $N = 3$, error bars are SD.

conversion of RAW files to mzXML[27]. Data was searched against the SwissProt human protein database using X! Tandem with the Orbitrap predefined method including reversed sequences[24].

Disulfide bond analysis

MS identification of disulfide bonds was carried out using methods to reduce bond scrambling[28]. Arthritic SF acidic fractions were analyzed in solution in reduced (4 M urea, 40 mM Tris, 25 mM DTT, pH 7, 70°C for 2 h, an additional 25 mM DTT added and incubated for a further 2 h) and non-reduced conditions. To inhibit bond scrambling, all samples were alkylated (a molar ratio of protein:4-vinylpyrridine of at least 1:6) then buffer exchanged with 40 mM Tris pH 7. Samples were then deglycosylated with 10 mU sialidase A (Prozyme Inc.) and 2.5 mU *O*-glycanase (Prozyme Inc.), 50 mM sodium phosphate pH 6, at 37°C for 16 h. Deglycosylated samples were digested in solution (pH 7) with lysine-C and trypsin as described above.

Intra-disulfide bonds and inter-disulfide bonds between COMP and lubricin were identified initially using MassMatrix software[25]. Parameters were set to the Orbitrap default method for disulfide bond identification with the addition of the S-pyridylethylation of cysteine modification. An abundant sample with a high proportion of large complexes (sample SpA1, Fig. 1(A)) was initially analyzed using a COMP–lubricin database (accession numbers P49747 and NP_005798) to identify disulfide bound peptides. The bonds identified via MassMatrix were then used to create specific bond databases, which were used to search all other samples to specifically target identifiable bonds. The MassMatrix default parameters for pp (5.0) and pp_{tag} (1.3) were used and all identified spectra were confirmed manually and only included in the data set when the combination of the *b*- and *y*-series provided almost complete coverage of all bound peptides.

Mass spectrometry methods and settings

All tryptic peptides were analyzed by nanoflow reverse-phase LC-electrospray ionization MS/MS using an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) as previously described[29]. Briefly, 3 μL of each digest was injected into a precolumn (4 cm × 100 μm inner diameter) and analytical column (20 cm × 50 μm inner diameter) packed with ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch, GmbH). A split flow rate of 100 nL/min was used. The gradient included 5 min of 100% A (aqueous 0.2% formic acid) followed by a 40 min gradient from 5 to 50% B (acetonitrile). Full MS scans were obtained in the Orbitrap at 400–2000 *m/z*, 2 microscans, maximum ion injection time 500 ms, and a target value of 500,000, using the lock mass feature for internal calibration (*m/z* 445.1200). Six data dependent MS/MS scans were acquired in the LTQ using collision induced fragmentation and a normalized collision energy of 30, activation energy of 0.250 and activation time of 30.

Statistical analyses

Solid-phase binding assays were performed in triplicate and repeated twice and data analyzed by Kruskal–Wallis test followed by Dunn's Multiple Comparison test of all pairs due to unequal variance, non-Gaussian distribution and sample size, using GraphPad Prism version 5. Proteomic and disulfide bond analyses were performed using X! Tandem[24] and MassMatrix[25] respectively, as described in the mass spectrometry methods section.

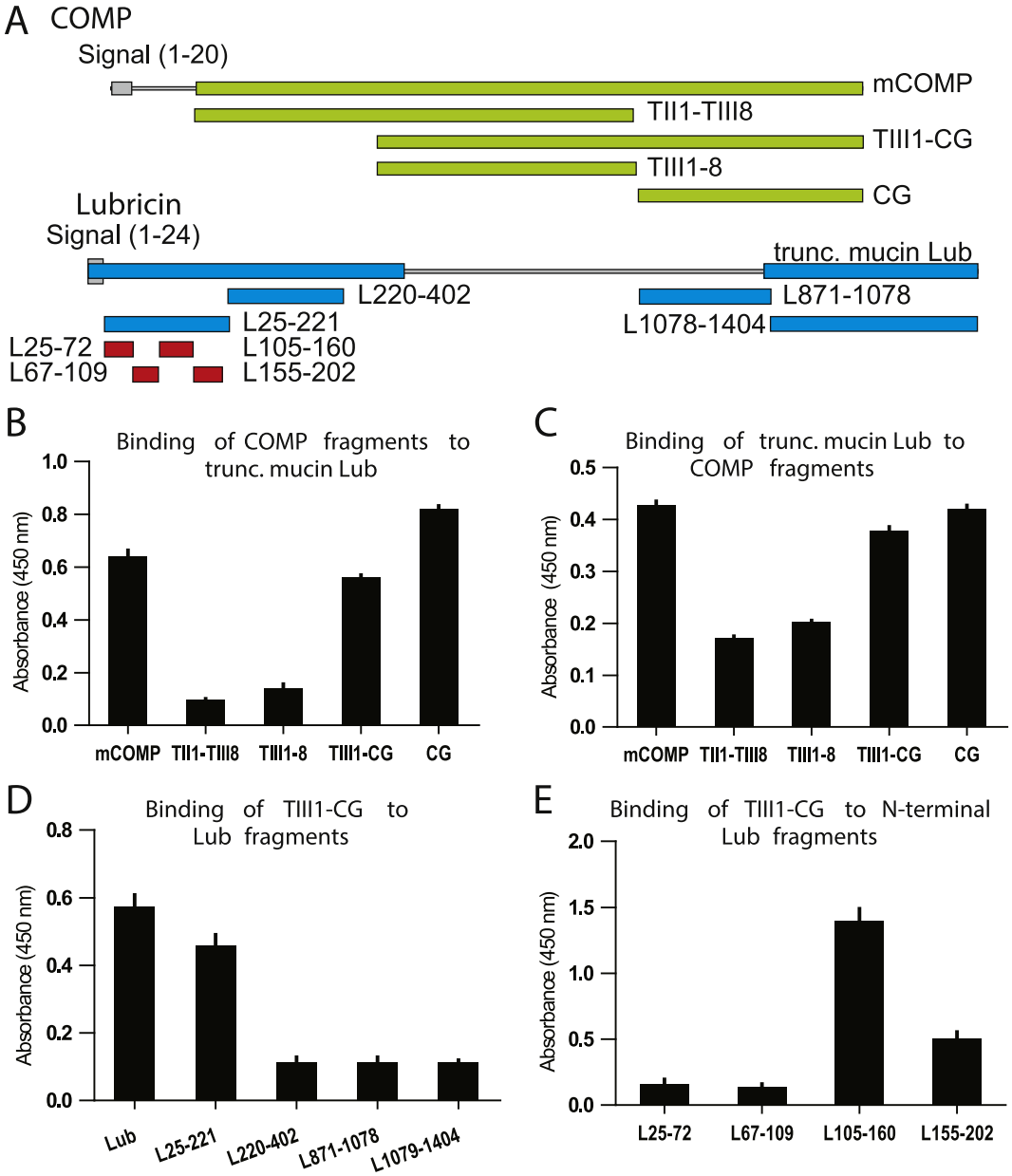

Fig. 2. RC COMP and lubricin form non-covalent complexes. (A) Representation of RC COMP and lubricin constructs. Green constructs: expressed, His-tagged, in mammalian 293-EBNA cells. Blue constructs: expressed, FLAG-tagged in mammalian 293F cells. Red constructs: expressed, GST-tagged in *E. coli* strain Rosetta 2. (B) Interaction between RC COMP fragments and truncated mucin domain, full-length lubricin by solid binding assay. RC lubricin was coated to ELISA plate. After blocking, His-tagged COMP fragments (mCOMP, TIII1-CG, TII1-TIII8, and TIII1-8) were added. Bound COMP was detected by anti-His antibody. $N = 3$, error bars are SD. Comparison of mCOMP to fragments showed no statistically significant difference by non-parametric test except mCOMP vs TII1-TIII8 ($p < 0.05$). (C) Binding trunc. mucin lubricin to COMP fragments. His-tagged COMP fragments (mCOMP, CG, TIII1-CG, and TII1-TIII8) were used to coat an ELISA plate. Trunc. mucin domain lubricin was added after blocking. Bound lubricin was detected by anti-FLAG antibody. $N = 3$, error bars are SD. Comparison of mCOMP to fragments showed no statistically significant difference by non-parametric test. (D) Binding RC COMP fragment (TIII1-CG) to lubricin fragments. RC trunc. mucin lubricin or fragments (L25-221, L220-402, L871-1078, and L1078-1404) were used to coat an ELISA plate. After blocking, COMP fragment (TIII1-CG) was added and detected by anti-His antibody. $N = 3$, error bars are SD. Comparison of Lub to fragments showed no statistically significant difference by non-parametric test. (E) Binding RC COMP fragment (TIII1-CG) to lubricin *N*-terminal fragments. RC lubricin fragments (L25-72, L67-109, L105-160 and L155-202) were used to coat an ELISA plate. After blocking, COMP fragment (TIII1-CG) was added, and detected by anti-His antibody. $N = 3$, error bars are SD. All pairs showed no statistically significant difference by non-parametric test except L67-109 vs L105-160 ($p < 0.05$).

Results

Discovery of COMP–lubricin complexes in arthritic SF

The SF acidic fraction from RA, OA and SpA patients ($n = 7$) were separated under non-reducing conditions and immunoblots probed with antibodies to COMP and lubricin (Fig. 1(A) and Fig. S1). The mono, di, tri, tetra and pentameric forms of COMP were prominent in the RA1 sample with the other samples containing less of the ladder. All samples contained lubricin. A range of very high molecular weight (HMW) bands contained both lubricin and COMP, suggesting the proteins were associating. Samples with a high amount of COMP oligomers contained low amounts of the potential COMP–lubricin complexes, and vice versa. This may relate to

disease state as the SpA sample contained more COMP–lubricin complex than the two RA samples. These HMW complexes were reducible with DTT as shown in Fig. S1.

The COMP and lubricin MS identifications from bands i-iv from patient SpA1 are shown in Fig. 1(B). Confidence of proteomic identifications is shown as an log-10 of the expectation value (log(e)), that is, the probability that the identification is a random assignment. Hence, the smaller the expectation value, the greater the confidence[30]. All identifications are of high certainty. The number of unique and total peptides used for each identification is also shown. Unique peptides and X! Tandem search results are listed in Fig. S2. Both COMP and lubricin were identified in all bands. Lubricin was most abundant in band i with a decrease in identified lubricin peptides with the lower MW bands. The opposite was observed with COMP, with the most peptides identified in band iv, suggesting that COMP and lubricin form a range of complexes with varied proportions of the two proteins.

Size exclusion chromatography was used to separate possible complexes in SF ($n = 2$). Fractions were then assessed by ELISA using anti-COMP mAb HC484D1 and anti-lubricin mAb13 antibodies. The chromatograms of samples with the lowest (RA1) and highest (SpA1) amounts of COMP–lubricin complexes are shown in Fig. 1(D) and (E) respectively. The RA1 sample has a dominant COMP-positive peak at a much lower MW (possibly the COMP pentamer) compared to the SpA sample. This may be due to the increased COMP–lubricin complexes in SpA1 as the major COMP peak corresponds to the higher MW lubricin peak.

To determine if the proteins were associated rather than co-migrating, two methods were used: co-immunoprecipitation (Co-IP) and sandwich ELISA. The Co-IP ($n = 2$) used anti-COMP mAb and protein G beads to capture COMP–lubricin complexes from the acidic fraction of SF. After a stringent wash (0.75% Triton X-100) and SDS-PAGE separation, Western blots were probed with anti-lubricin mAb13 to identify only COMP–lubricin complexes. The complex was enriched in SF from RA and SpA patients [Fig. 1(C)]. The sandwich ELISA ($n = 1$ shown) employed rat anti-COMP mAb to capture the complexes and anti-lubricin monoclonal antibodies for detection. This detected COMP–lubricin complexes in SF from an RA patient in a dose-dependent manner [Fig. 1(F)]. These data indicate that lubricin and COMP are associated in RA, OA (Sup. S1) and SpA SF.

Non-covalent interactions between RC versions of COMP and lubricin

To identify regions of non-covalent interactions between COMP and lubricin, RC fragments of the proteins were expressed [Fig. 2(A)]. To increase expression yields, the mucin-like domain of lubricin was shortened by the removal of AA 403-870. Apparent MW and purity (>95%) were verified for each RC protein (see Fig. S3).

Fragment interactions were tested by solid phase assays; truncated mucin-domain full-length lubricin (trun. mucin lubricin) was conjugated to microtitre plates and incubated with various COMP fragments that together cover the full sequence of COMP, except the *N*-terminal coiled-coil domain [Fig. 2(B)]. The terminal COMP fragment TIII1-CG, encompassing the TSP type III repeats and the *C*-terminal globule (AA 268-757), interacted with lubricin, as did the CG domain alone. A similar result was obtained by a reversed assay where the plate was coated with COMP fragments and probed with trunc. mucin lubricin. This narrowed the COMP interaction area to within the *C*-terminal globule (CG, AA 518-757), given that the TII1-III8 (AA 73-517) fragment did not associate with the trunc. mucin lubricin and that all fragments containing the *C*-terminal globule did [Fig. 2(C)].

The area of interaction within lubricin was also addressed. Lubricin fragments were coated onto microtitre plates and incubated with the TIII1-CG (AA 268-757) fragment containing the COMP binding region AA 518-757 [Fig. 2(D)]. Using mammalian expressed fragments, the *N*-terminal domain (L25-221), proximal to the mucin-like domain [Fig. 2(D)] had the highest COMP affinity, apart from the truncated mucin-full length lubricin construct, which was slightly higher. To define the interacting area, smaller *N*-terminal lubricin fragments (L25-72, L67-109, L105-160 and L155-202) were expressed in *E. coli* given that this region of lubricin is not highly glycosylated. L105-160 had the strongest affinity, with L25-72 and L67-109 showing almost no binding to COMP [Fig. 2(E)].

Altogether, these *in vitro* results demonstrate that the *C*-terminal domain of COMP (AA 518-757) spontaneously interacts non-covalently with the *N*-terminal of lubricin (AA 105-202). As the lubricin fragment was expressed in *E. coli*, this interaction is protein mediated. The COMP fragments, on the other hand, were expressed in mammalian cells, so post-translational modification involvement in COMP bonding cannot be excluded.

COMP and lubricin form a network of protein complexes stabilized by disulfide bonds

As the COMP–lubricin complex was identified in highly dissociating conditions, arthritic SF acidic fractions were analyzed for disulfide bonds by liquid chromatography (LC)-MS/MS. Non-reduced and reduced (negative control) samples were all treated with sialidase and *O*-glycanase to allow for more effective trypsin digestion. All samples were alkylated and preparation done at or below pH 7 to prevent bond scrambling. Sample SpA1 was chosen for initial analysis given its high abundance of COMP–lubricin complexes and large sample size. Bonds identified in this sample were compared to other samples with lower complex abundance. Bonds were identified using MassMatrix software[27] and the MS/MS spectra used for identification were confirmed manually. An example of MS/MS spectra identifying disulfide bonds is shown in supplementary Fig. S4 and all bonds identified are shown in supplementary Fig. S5.

The location of disulfide bonds within the protein domain structures identified in sample SpA1 are shown in Fig. 3(A). Bond forming cysteines of lubricin are focused to the *N*-terminal, with 13 of the 17 inter-protein bonds identified between AA 64 and 86 [Fig. 3(A)]. All other bond forming cysteines were located in the *C*-terminal hemopexin domain [Fig. 4(D)], except for the single cysteine within the mucin domain (MUC) of lubricin [Fig. 4(C)]. The bonds identified in sample RA1 are also shown in Fig. 3(B). Bonds consistent between these two samples are focused in the *N*-terminal lubricin domain (Lub AA 64-86), indicating their importance for complex formation. The inter-disulfide bond-forming cysteines of COMP are not as highly concentrated, extending throughout COMP except for the TSP *C*-terminal domain. The region COMP AA 102-191 (corresponding to the EGF/TSP type II repeats) as well as the region COMP AA 287-504 (corresponding to the TSP type III repeats) contain all of the inter-protein disulfide bond-forming cysteines [Fig. 3(B)].

Analysis of the reduced SpA1 sample identified that two disulfide bonds (lubricin C70 and COMP C371, and lubricin C64 and COMP C328) were still present. A single bond (lubricin C70 and COMP C371) was identified in both reduced samples (RA1 and SpA1). As the large protein complexes formed may not be fully-reducible, this suggests these bonds at the *N*-terminal of lubricin and *C*-terminal of COMP are important to complex formation and difficult to reduce. The consistent identification of complex tri-peptides, rather than di-peptide complexes (Fig. 4) shows that COMP and lubricin form a complex network.

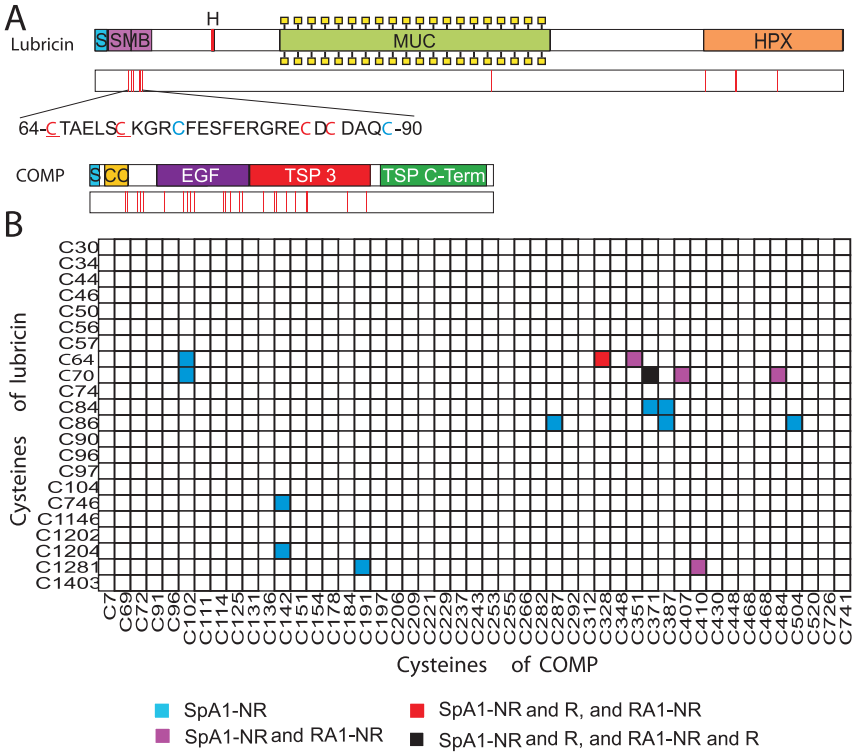

Fig. 3. Location of COMP and lubricin inter-protein disulfide bonds. (A) Location of cysteines involved in disulfide bonds within COMP and lubricin. Domain structures shown for COMP and lubricin. Lubricin's signal peptide (S), 2× somatomedin-B domains (SMB), heparin binding domain (H), mucin domain (MUC), hemopexin repeats (HPX). COMP's signal peptide (S), coiled coil domain (CC), 4× EGF Ca-binding domains (EGF), TSP type 3 repeats (TSP 3), TSP *C*-terminal domain (TSP *C*-Term). Red lines: cysteines involved in disulfide bonds. The lubricin sequence between AA 64 and 90 shown, red: cysteines involved in inter-protein disulfide bonds, blue: free cysteines in native lubricin (detected as alkylated or free by MS in non-reduced (NR) samples). (B) Inter-protein disulfide bonds from two samples SpA1 (reduced (R) and non-reduced (NR)) and RA1 (R and NR). Cysteines of lubricin are shown on *y*-axis and COMP on *x*-axis, bonds between them by a colored square. Blue: bonds only identified in SpA1-NR, Purple in SpA1-NR and RA1-NR, Red in SpA-NR and R as well as RA1-NR and black in SpA-NR and R and RA1-NR and R.

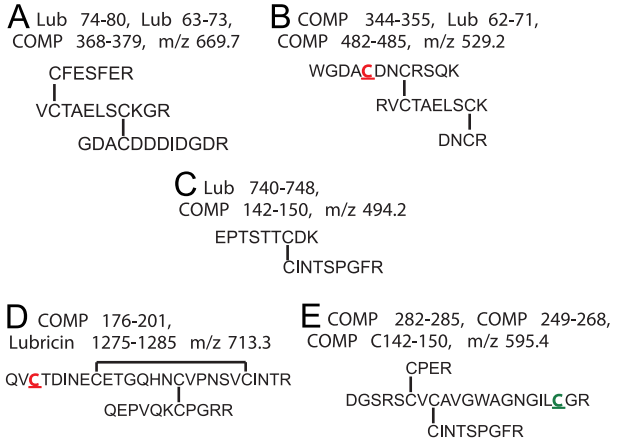

Fig. 4. Types of disulfide bonds identified. Examples of the peptides bound by disulfide bonds including di- and tri-peptide complexes. The peptides' AA range is shown above the bound peptides along with the *m*/*z*. Bonds are shown by black lines. Red (detected as alkylated) and green (detected without alkylation) underlined were present as free cysteines in native SF lubricin and COMP.

Alkylated and free cysteines were identified in non-reduced samples (Fig. 4(B), (D) and (E) and S5), showing that both proteins retain free cysteines. In the *N*-terminal of lubricin, cysteine C90 was identified as a free cysteine, suggesting this domain adjacent to the lubricin disulfide binding region (AA 64-86) remains free, enabling it to form complexes in the ECM. The MS analysis confirms that the COMP–lubricin complex is stabilized by a large, complex network of disulfide bonds. It also shows that the *N*-terminal disulfide binding region of lubricin (Lub AA 60-86) is highly important for complex formation, while COMP cysteines in a larger region of the protein (COMP AA 102-191 and COMP AA 287-504) are involved.

Intra-protein disulfide bonds

Intra-protein bonds were also identified in lubricin and COMP (Fig. 5). COMP intra-protein bonds ($n = 11$) (intra-peptide and inter-peptide) were evenly distributed throughout the protein. Two disulfide inter-peptide bonds were located between C91–C96 and C184–C197. The intra-peptide COMP bond between C176 and C201 [Fig. 4(D)] has previously been suggested. The confirmation of this bond validates our approach, but also highlights that some of the additional previously reported disulfide bonds[31] were not identified. This is likely due to the difficulty in digesting such large protein complexes in a non-reduced state. In lubricin, six intra-protein bonds were identified. Of these, three involved the C64 and/or C70 residues and four are involved in disulfide bound tripeptides that also include inter-protein bonds to COMP. Lubricin intra-protein bonds, C746–C1146 and C1204–C1281 were the only intra-peptides identified solely containing lubricin peptides. The analysis is unable to define if these bonds are within a single protein molecule and therefore involved in secondary structure formation or if these bonds are between two separate molecules and are the bonds involved in the proposed lubricin dimer formation[32].

Overall these data show that COMP and lubricin form a complex disulfide-bonded network. The identification of bound COMP and

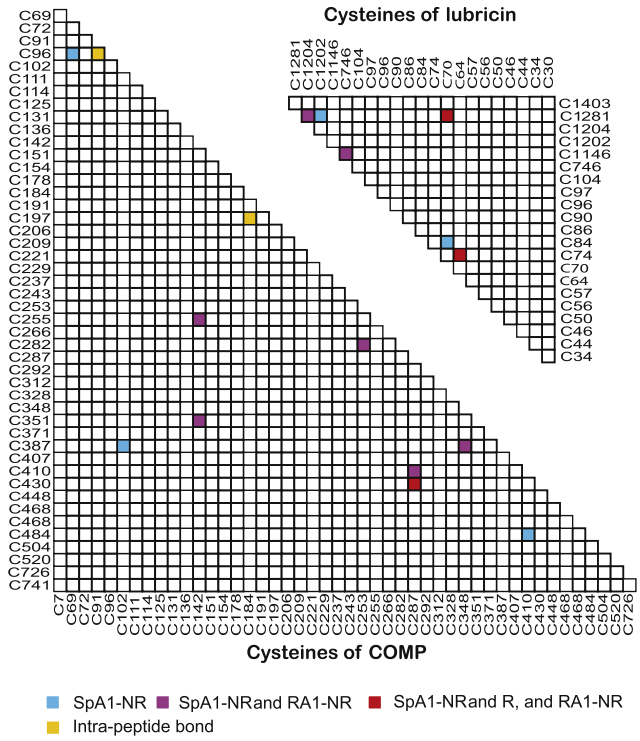

Fig. 5. Intra-protein disulfide bonds identified in lubricin and COMP. Intra-protein disulfide bonds (inter- and intra-peptides) identified in two samples, SpA1 (reduced (R) and non-reduced (NR)) and RA1 (R and NR). Cysteines from the same protein are on both axes, bonds between them shown by a colored square. Blue: bonds only identified in SpA1-NR, Purple in SpA1-NR and RA1-NR, and Red in SpA-NR and R as well as RA1-NR. Intra-peptide bonds in yellow. In lubricin, 6 intra-protein bonds were identified; 3 involved the C64 or C70 residues. COMP intra-protein bonds (n = 11) were evenly distributed throughout the protein.

lubricin molecules in the SF of arthritis patients suggests that this complex may be localized to the outer cartilage layer and lost into the SF in disease.

Discussion

We determined here that COMP and lubricin form non-covalently and disulfide bound complexes identifiable in the SF of RA, OA and SpA patients. Using truncated RC proteins and mass spectrometric techniques we confirmed this specific interaction.

COMP is an integral structural component of cartilage, assembling and stabilizing the ECM through interactions with other ECM proteins, aided by its flexible homopentameric structure[19,33]. The *C*-terminal of COMP (AA 579-595 of 757) binds to fibronectin[17], aggrecan[16] and cartilage collagens I, II and IX[34–36]. RC studies completed here showed lubricin also non-covalently binds to the *C*-terminal of COMP (AA 518-757). The *C*-terminal forms the pistil of the COMP protein flower-like structure while the TSP type III repeats (AA 268-528) forms the petals[37], shown here to form covalent bonds with lubricin. The pistil head and petal are separated by 35–50 Å (PDB 3FBY), similar to the distance between the two binding regions of lubricin. Other proposed covalent linking sites may indicate crosslinking of lubricin between COMP pentamers. This suggests a broader range of binding potential than previously shown for COMP.

The binding capabilities of lubricin are less well known, however bovine synovial lubricin multimers have been identified[38]. Interactions between lubricin and denatured, amorphous and fibrilliar chicken type II collagen have been identified, although the mechanism is unknown[18]. Here, the non-covalent binding region of lubricin was localized to the *N*-terminal (AA 105-202 of 1404), a protein mediated interaction as the lubricin fragment was expressed in *E. coli*. The disulfide binding region (AA 64-86) was also identified in the *N*-terminal somatomedin B domain of lubricin. The two binding regions are directly adjacent, suggesting a two stage mechanism where lubricin first binds COMP via a non-covalent bond to allow the formation of the stronger disulfide bonds to strengthen the complex. This mechanism may be necessary due to the excessive mechanical strain in the moving joint.

Disulfide bond creation is not limited to the endoplasmic reticulum; quiescin sulfhydryl oxidase 1 (QSOX1) secreted by quiescent fibroblasts is active in disulfide bond formation in the ECM[39,40]. QSOX1 is necessary for the incorporation of laminin into the ECM[41], and EST frequencies suggest high expression in connective tissue, with even greater levels in chondrosarcoma where cartilage component production is high[40]. Free cysteines on extracellular COMP and lubricin are also necessary for disulfide bond formation, and were identified by MS analysis in both proteins. The previous identification of a free cysteine residue in the globular fold (AA 726) by the crystal structure of the COMP *C*-terminal (AA 225-757)[37], supports our detection of COMP free cysteines. The identification of the *N*-terminal disulfide binding region of lubricin as a concentration of disulfide-forming cysteines suggests this may be a protected region. Interestingly, detailed site-specific glycosylation analysis of lubricin showed no glycosylation on the initial 130 AA despite *O*-glycosylation being possible in this region[42]. The binding region (AA 64-86) is glycosylation free, allowing disulfide bond formation. Also, the initiation of glycosylation directly after this region may protect free cysteine residues from further reactions, allowing secretion of cysteine –SH side chains into the ECM.

The COMP–lubricin complex was identified in arthritic SF, however, the concentration of COMP is far greater in cartilage than other joint tissues, including the SF[43] suggesting the complexes originate from the cartilage. To be lost into the SF in degenerative arthritis, as observed by increased matrix metalloproteinase-3 degradation of ECM components in OA[44] would require the presence of the complex at the cartilage surface. Therefore, these data could suggest that the COMP–lubricin complex is produced in the cartilage, localized on the cartilage surface and is lost into the SF during disease. Further analysis on the localization of these complexes in the joint is essential to thoroughly understand their significance.

Lubricin has begun to represent a possible biologic for the treatment of arthritic diseases where the lubricative surface has been lost, or even to prevent damage[5,45]. Lubricin injected into a damaged joint localizes to damaged surfaces[46], maybe due to the exposure of COMP during cartilage degradation. Lubricin retention in the rat joint has shown RC lubricin detectable after 4 weeks[46]. The increased understanding of lubricin attachment from this study may allow the design of a longer lasting lubricative surface, perhaps by constructing RC lubricin with additional COMP binding zones or the addition of COMP–lubricin complexes to improve incorporation into the cartilage.

Contributions

Conception and design (SF, SK, NK and CJ). Analysis and interpretation of data (SF, SK, LA, JRR, NK and CJ). Drafting the article (SF and NK). All authors critically reviewed and approved the manuscript for submission. Provision of study materials and associated expertise (LB and AA). Statistical expertise (SF). Obtaining funding (NK and SK). Collection and assembly of data (SF and NK). Niclas Karlsson (niclas.karlsson@medkem.gu.se) and Sarah Flowers

(sarah.flowers@medkem.gu.se) take responsibility for the integrity of the work as a whole.

Competing interests

All funding sources are stated in funding section. SF, SK, NK and CJ have submitted a patent application related to this work.

Funding and role of the funding sources

Funding sources had no role in the study design, collection, analysis or interpretation of data; writing of the manuscript or the decision to submit the manuscript for publication. NK, SF, CJ, JRR and LA were supported by The Swedish Foundation for International Cooperation in Research and Higher Education (STINT), Kung Gustav V:s 80-års foundation, Petrus and Augusta Hedlund's foundation and AFA insurance research fund. The LTQ-Orbitrap mass spectrometer was obtained by a grant from the Knut and Alice Wallenberg Foundation (KAW2007.0118). HPLC- instrument was obtained with a gift from the Ingabritt and Arne Lundbergs Research Foundation. SK and AA acknowledge funding from the Crafoord Foundation, Alfred Österlund Foundation and Greta and Johan Kock Foundation.

Acknowledgements

We thank C. R. Flannery (Pfizer, Cambridge, MA) for providing lubricin-specific antibody (mAb 13). To the lasting memory of Dick Heinegård and his lifetime of dedication to science in Sweden.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.joca.2017.03.016.

References

1. Lee AS, Ellman MB, Yan D, Kroin JS, Cole BJ, van Wijnen AJ, *et al*. A current review of molecular mechanisms regarding osteoarthritis and pain. Gene 2013;527:440–7.
2. Firestein GS. Evolving concepts of rheumatoid arthritis. Nature 2003;423:356–61.
3. Rosman Z, Shoenfeld Y, Zandman-Goddard G. Biologic therapy for autoimmune diseases: an update. BMC Med 2013;11:88.
4. Gleghorn JP, Bonassar LJ. Lubrication mode analysis of articular cartilage using Stribeck surfaces. J Biomech 2008;41:1910–8.
5. Waller KA, Zhang LX, Elsaid KA, Fleming BC, Warman ML, Jay GD. Role of lubricin and boundary lubrication in the prevention of chondrocyte apoptosis. Proc Natl Acad Sci U S A 2013;110:5852–7.
6. Jones AR, Gleghorn JP, Hughes CE, Fitz LJ, Zollner R, Wainwright SD, *et al*. Binding and localization of recombinant lubricin to articular cartilage surfaces. J Orthop Res 2007;25:283–92.
7. Nugent-Derfus GE, Chan AH, Schumacher BL, Sah RL. PRG4 exchange between the articular cartilage surface and synovial fluid. J Orthop Res 2007;25:1269–76.
8. Su JL, Schumacher BL, Lindley KM, Soloveychik V, Burkhart W, Triantafillou JA, *et al*. Detection of superficial zone protein in human and animal body fluids by cross-species monoclonal antibodies specific to superficial zone protein. Hybridoma 2001;20:149–57.
9. Liu YJ, Lu SH, Xu B, Yang RC, Ren Q, Liu B, *et al*. Hemangiopoietin, a novel human growth factor for the primitive cells of both hematopoietic and endothelial cell lineages. Blood 2004;103:4449–56.
10. Schmidt TA, Sullivan DA, Knop E, Richards SM, Knop N, Liu S, *et al*. Transcription, translation, and function of lubricin, a boundary lubricant, at the ocular surface. JAMA Ophthalmol 2013;131:766–76.
11. Flannery CR, Hughes CE, Schumacher BL, Tudor D, Aydelotte MB, Kuettner KE, *et al*. Articular cartilage superficial zone protein (SZP) is homologous to megakaryocyte stimulating factor precursor and Is a multifunctional proteoglycan with potential growth-promoting, cytoprotective, and lubricating properties in cartilage metabolism. Biochem Biophys Res Commun 1999;254:535–41.
12. Rees SG, Davies JR, Tudor D, Flannery CR, Hughes CE, Dent CM, *et al*. Immunolocalisation and expression of proteoglycan 4 (cartilage superficial zone proteoglycan) in tendon. Matrix Biol 2002;21:593–602.
13. Estrella RP, Whitelock JM, Packer NH, Karlsson NG. The glycosylation of human synovial lubricin: implications for its role in inflammation. Biochem J 2010;429:359–67.
14. Ali L, Jin C, Karlsson NG. Glycoproteomics of lubricin-implication of important biological glyco- and peptide-epitopes in synovial fluid. In: Lemmey AB, Ed. Rheumatoid arthritis-etiology, consequences and co-morbidities. InTech; 2012.
15. Oldberg A, Antonsson P, Lindblom K, Heinegard D. COMP (cartilage oligomeric matrix protein) is structurally related to the thrombospondins. J Biol Chem 1992;267:22346–50.
16. Chen FH, Herndon ME, Patel N, Hecht JT, Tuan RS, Lawler J. Interaction of cartilage oligomeric matrix protein/thrombospondin 5 with aggrecan. J Biol Chem 2007;282:24591–8.
17. Di Cesare PE, Chen FS, Moergelin M, Carlson CS, Leslie MP, Perris R, *et al*. Matrix-matrix interaction of cartilage oligomeric matrix protein and fibronectin. Matrix Biol 2002;21:461–70.
18. Chang DP, Guilak F, Jay GD, Zauscher S. Interaction of lubricin with type II collagen surfaces: adsorption, friction, and normal forces. J Biomech 2014;47:659–66.
19. Acharya C, Yik JH, Kishore A, Van Dinh V, Di Cesare PE, Haudenschild DR. Cartilage oligomeric matrix protein and its binding partners in the cartilage extracellular matrix: interaction, regulation and role in chondrogenesis. Matrix Biol 2014;37:102–11.
20. Mann HH, Ozbek S, Engel J, Paulsson M, Wagener R. Interactions between the cartilage oligomeric matrix protein and matrilins. Implications for matrix assembly and the pathogenesis of chondrodysplasias. J Biol Chem 2004;279:25294–8.
21. Happonen KE, Saxne T, Aspberg A, Morgelin M, Heinegard D, Blom AM. Regulation of complement by cartilage oligomeric matrix protein allows for a novel molecular diagnostic principle in rheumatoid arthritis. Arthritis Rheum 2010;62:3574–83.
22. Verma P, Dalal K. Serum cartilage oligomeric matrix protein (COMP) in knee osteoarthritis: a novel diagnostic and prognostic biomarker. J Orthop Res 2013;31:999–1006.
23. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
24. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 2004;20:1466–7.
25. Xu H, Zhang L, Freitas MA. Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. J Proteome Res 2008;7:138–44.
26. Abramoff MD, Magalhaes PJ, Ram SJ. Image processing with ImageJ. Biophot Int 2004;11:36–42.

27. Xu H, Freitas MA. Automated diagnosis of LC-MS/MS performance. Bioinformatics 2009;25:1341–3.
28. Go EP, Zhang Y, Menon S, Desaire H. Analysis of the disulfide bond arrangement of the HIV-1 envelope protein CON-S gp140 DeltaCFI shows variability in the V1 and V2 regions. J Proteome Res 2011;10:578–91.
29. Andersch-Bjorkman Y, Thomsson KA, Holmen Larsson JM, Ekerhovd E, Hansson GC. Large scale identification of proteins, mucins, and their O-glycosylation in the endocervical mucus during the menstrual cycle. Mol Cell Proteomics 2007;6:708–16.
30. Eriksson J, Fenyo D. Probity: a protein identification algorithm with accurate assignment of the statistical significance of the results. J Proteome Res 2004;3:32–6.
31. Malashkevich VN, Kammerer RA, Efimov VP, Schulthess T, Engel J. The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? Science 1996;274:761–5.
32. Jay GD, Waller KA. The biology of lubricin: near frictionless joint motion. Matrix Biol 2014;39:17–24.
33. Haleem-Smith H, Calderon R, Song Y, Tuan RS, Chen FH. Cartilage oligomeric matrix protein enhances matrix assembly during chondrogenesis of human mesenchymal stem cells. J Cell Biochem 2012;113:1245–52.
34. Holden P, Meadows RS, Chapman KL, Grant ME, Kadler KE, Briggs MD. Cartilage oligomeric matrix protein interacts with type IX collagen, and disruptions to these interactions identify a pathogenetic mechanism in a bone dysplasia family. J Biol Chem 2001;276:6046–55.
35. Rosenberg K, Olsson H, Morgelin M, Heinegard D. Cartilage oligomeric matrix protein shows high affinity zinc-dependent interaction with triple helical collagen. J Biol Chem 1998;273:20397–403.
36. Vranka J, Mokashi A, Keene DR, Tufa S, Corson G, Sussman M, *et al.* Selective intracellular retention of extracellular matrix proteins and chaperones associated with pseudoachondroplasia. Matrix Biol 2001;20:439–50.
37. Tan K, Duquette M, Joachimiak A, Lawler J. The crystal structure of the signature domain of cartilage oligomeric matrix protein: implications for collagen, glycosaminoglycan and integrin binding. FASEB J 2009;23:2490–501.
38. Schmidt TA, Plaas AH, Sandy JD. Disulfide-bonded multimers of proteoglycan 4 PRG4 are present in normal synovial fluids. Biochim Biophys Acta 2009;1790:375–84.
39. Alon A, Grossman I, Gat Y, Kodali VK, DiMaio F, Mehlman T, *et al.* The dynamic disulphide relay of quiescin sulphydryl oxidase. Nature 2012;488:414–8.
40. Kodali VK, Thorpe C. Oxidative protein folding and the Quiescin-sulfhydryl oxidase family of flavoproteins. Antioxidants Redox Signal 2010;13:1217–30.
41. Ilani T, Alon A, Grossman I, Horowitz B, Kartvelishvily E, Cohen SR, *et al.* A secreted disulfide catalyst controls extracellular matrix composition and function. Science 2013;341:74–6.
42. Ali L, Flowers SA, Jin C, Bennet EP, Ekwall AK, Karlsson NG. The O-glycomap of lubricin, a novel mucin responsible for joint lubrication, identified by site-specific glycopeptide analysis. Mol Cell Proteomics 2014;13:3396–409.
43. Muller G, Michel A, Altenburg E. COMP (cartilage oligomeric matrix protein) is synthesized in ligament, tendon, meniscus, and articular cartilage. Connect Tissue Res 1998;39:233–44.
44. Berenbaum F. Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). Osteoarthr Cartil 2013;21:16–21.
45. Jay GD, Fleming BC, Watkins BA, McHugh KA, Anderson SC, Zhang LX, *et al.* Prevention of cartilage degeneration and restoration of chondroprotection by lubricin tribosupplementation in the rat following anterior cruciate ligament transection. Arthritis Rheum 2010;62:2382–91.
46. Vugmeyster Y, Wang Q, Xu X, Harrold J, Daugusta D, Li J, *et al.* Disposition of human recombinant lubricin in naive rats and in a rat model of post-traumatic arthritis after intra-articular or intravenous administration. AAPS J 2012;14:97–104.