COLLAGEN VI: CONFORMATION OF A-DOMAIN ARRAYS AND MICROFIBRIL ARCHITECTURE
Nicola Beecher1*, Alan M. Roseman2*, Thomas A. Jowitt1, Richard Berry1, Helen Troilo1, Richard A. Kammerer1,3, C. Adrian Shuttleworth1, Cay M. Kielty1, Clair Baldock1.

From the 1Wellcome Trust Centre for Cell-Matrix Research, 2Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK. 3Now at Laboratory of Biomolecular Research, OFLC/106, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland.

Running head: Collagen VI conformation and architecture
* Both authors contributed equally to this work.

Address correspondence to: Clair Baldock, Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK. Fax 0161 275 5082; E-mail clair.baldock@manchester.ac.uk

Collagen VI is a ubiquitous extracellular matrix protein that assembles into beaded microfibrils that form networks linking cells to the matrix. Collagen VI microfibrils are typically formed from a heterotrimer of the α1, α2 and α3 chains. The α3 chain is distinct as it contains an extended N-terminus with up to ten consecutive von Willebrand factor type A-domains (VWA). Here we use solution small angle X-ray scattering (SAXS) and single particle analysis EM to determine the nanostructure of nine of these contiguous A-domains. Both techniques reveal a tight C-shape conformation for the A-domains. Further, using biophysical approaches we demonstrate that the N-terminal region undergoes a conformational change and a proportion forms dimers in the presence of Zn2+. This is the first indication that divalent cations interact with collagen VI A-domains. A 3D reconstruction of tissue-purified collagen VI microfibrils was generated using EM and single particle image analysis. The reconstruction showed the intricate architecture of the collagen VI globular regions, in particular the highly structurally conserved C-terminal region, and variations in the appearance of the N-terminal region. The N-terminal domains project out from the globular beaded region like angled radial spokes. These could potentially provide interactive surfaces for other cell-matrix molecules.

Collagen VI is a microfibrillar collagen with a distinctive ‘beads on a string’ appearance that is found in the extracellular matrix (ECM) of virtually all connective tissues. Until recently, it was thought to be composed of just three genetically distinct α-chains, α1(VI), α2(VI) and α3(VI), each comprising a short collagenous region with N- and C-terminal globular domains. Three additional collagen VI chains were recently discovered, α4(VI), α5(VI) and α6(VI), in both mice (1) and humans (2). The new chains have homology to the α3(VI) chain and the α6(VI) chain is widely expressed in a range of fetal and adult tissues (2).

Despite the widespread distribution of collagen VI, mutations in the collagen VI genes manifest almost purely as musculoskeletal in nature, as highlighted by the genetic linkage between defects in collagen VI and the heritable conditions Ulrich congenital muscular dystrophy (UCMD) and Bethlem myopathy (for review see (3)). Bethlem myopathy is at the mild end of the clinical spectrum and is associated with reduced collagen VI levels, whereas UCMD is a potentially fatal disorder where a complete lack of collagen VI can occur (4).

Collagen VI has a hierarchical assembly, the first stage of which is the formation of a triple-helical monomer. Next the monomers associate intracellularly to form dimers, then assemble into tetramers (5). The tetramers are secreted into the ECM and form microfibrils via non-covalent interactions (6,7). The majority of the globular domains of collagen VI show homology to VWA domains (8). The α3(VI) chain is intriguing because, whereas the α1(VI) and α2(VI) chains are similar in size and contain one N- and two C-terminal domains, the α3(VI) chain is much larger with a maximum of ten N-terminal A-domains that are subject to alternate splicing, and also C-terminal domains akin to type III fibronectin repeats and Kunitz domains (9,10). The α4(VI), α5(VI) and α6(VI) chains also have extensive globular N-termini (2). A-domains are found in a
range of ECM proteins including integrins, collagens and matrilins. In some cases, they have a metal-ion dependent adhesion site (MIDAS), that in the case of integrins, is required for Mn\(^{2+}\) or Mg\(^{2+}\)-dependent ligand binding (11).

The C-terminal domains of collagen VI are critical for assembly (12) and microfibril formation (13,14). It has also been established that the presence of at least five of the N-terminal α3 A-domains (N5-N1) are essential for microfibril formation (15). It is also likely that the additional N-terminal domains of the α3 chain afford the collagen VI microfibril the ability to participate in matrix interactions. The interaction of collagen VI with matrix molecules such as heparin/heparan sulphate, hyaluronan (16), the transforming growth factor-β induced gene-h3 (βig-h3) (17), biglycan, decorin (19) and matrilin-1 (20) signifies that collagen VI provides a physical link between ECM components and connective tissue cells. Thus, it is important to elucidate the role of these additional domains in order to shed light on both collagen VI-ECM interactions and the processes involved in microfibril formation.

To examine the role of the N-terminal region of the α3(VI) chain and to gain a clearer perspective of its function in microfibril formation, structural studies were performed on the human α3(VI) N9-N1 region. We have utilised Multiangle Laser Light Scattering (MALLS), analytical ultracentrifugation (AUC), EM with single particle analysis and SAXS to investigate the structure of this region. In addition, we studied the effect of metal ions on the conformation and oligomeric state of this region. To investigate the structure and conformation of the N-terminal region in tissue microfibrils we used EM and single particle analysis to generate a 3D reconstruction. The 3D microfibril structure reveals a high level of structural homogeneity in the C-terminal regions but a degree of conformational flexibility in the N-terminal regions.

**Experimental procedures**

*Expression and purification of the human α3(VI) N-terminal N9-N1 region* - The α3(VI) N9-N1 region (amino acids 237 - 2028) was generated by PCR from the construct described in (16) (gift from Professor M.L. Chu). A 10x histidine tag was incorporated at the C-terminus following a thrombin cleavage site. The construct was ligated into a modified pCEP-Pu vector (21) and transfected into HEK 293-EBNA cells cultured as described previously (22). The recombinant protein was purified by nickel affinity chromatography after which the histidine-tag was removed by thrombin digestion. Thrombin (1unit/mg) was incubated with the N9-N1 region in 50 mM Tris-HCl (pH 7.4) containing 150 mM NaCl at 20 °C overnight. Following thrombin digestion, the sample was passed over the nickel affinity column again, in order to remove any remaining tagged protein or tag. Size-exclusion chromatography on an AKTA purifier HPLC using a Superdex200 10/300GL column (GE Healthcare) in 50 mM Tris-HCl (pH 7.4) containing 150 mM NaCl was used as a final purification step to remove thrombin. Where needed, the protein was concentrated using Vivaspin centrifugal concentrators (Sartorius). The protein identity was confirmed by in-gel trypsin digestion and liquid chromatography tandem mass spectrometry (LC-MS/MS) using a NanoAcquity LC (Waters) coupled to a LTQ Velos (Thermo Fisher Scientific).

*MALLS Analysis* - Samples (0.5 ml at approximately 0.5 mg/ml) were loaded onto a Superdex200 10/300GL column running at a flow rate of 0.5 ml/min in 50mM Tris-HCl (pH 7.4) containing 150 mM NaCl. Samples eluting from the column passed through a DAWN Wyatt EOS 18-angle laser photometer. This was coupled to a Wyatt Optilab rEX refractive index detector and the molecular mass and concentrations of the resulting peaks were analysed using Astra 5.3.2.

*SAXS Solution* - Data on the N9-N1 region at ~5 mg/ml were collected at station 2.1 of the Synchrotron Radiation Source at Daresbury Laboratory, UK. The sample was maintained at 10 °C during data collection. Images were collected in multiple 60 sec frames at 1 m and 4 m sample-to-detector distances and the resulting profiles merged to cover a momentum transfer interval of 0.012 Å\(^{-1}\) < q < 0.8 Å\(^{-1}\). The forward scattering intensity, Rg and one-dimensional intra-particle distance distribution function p(r) in real space were evaluated with the indirect Fourier transform program GNOM (23) and particle shapes were restored ab initio using DAMMIN (24). Multiple
runs were performed to generate 20 similar shapes that were then combined and filtered to produce an averaged model using the DAMAVER (25) software package. Rigid body modelling to the experimental scattering data was performed using SASREF (26) with the C2 α2(VI) domain as the starting model (12), and domain linker lengths of 0.8 - 2.5 nm were tested. A 1.5 nm distance between consecutive domains gave the best fit (chi value = 2.1) to the experimental data and this distance constraint was used for 10 repeat simulations.

**EM and Single-Particle Analysis**- The N9-N1 region (40 µg/ml) was adsorbed onto glow-discharged carbon-coated grids and stained with 4% (w/v) uranyl acetate (pH 4.7). Grids were observed using an FEI Tecnai Twin transmission EM operating at 120 kV. Images were recorded under low dose conditions (<10 e/Å²) on a 2,048 x 2,048 pixel CCD camera at 52,000x magnification between -0.5 and -1.0 µm defocus. Images were converted to Imagic5 format (Image Science, Berlin); the Imagic5 suite of programs was used for particle picking and image processing (27). The total number of particles in the dataset was 1,776. Selected particles were band-pass-filtered with a high frequency cut-off of 20 Å and a low-frequency cut-off of 170 Å. Characteristic class-sum images were used as references to align the dataset during iterative rounds of translational and rotational alignment. Symmetry was not applied at any stage of the image analysis to the dataset. Euler angles were assigned to class-sum images, enabling calculation of an initial 3D reconstruction which was then subjected to multiple rounds of iterative refinement.

**Farfield methods**- Dual polarization interferometry was performed using an Analight 4D instrument from Farfield using 20 mM HEPES, 150 mM NaCl, pH7.4 as the running buffer. Following calibration of an amine functionalized sensor chip type FB80, lane 1 was activated by injecting 2 mg/ml BS3 cross-linker at 25 µl/min. Following this 200 µl of 100 µg/ml collagen VI was injected again at 25 µl/min. After stabilization there were successive injections of 500 µM MgCl₂, MnCl₂ and ZnCl₂ then ZnCl₂ between 25 and 500 µM followed each time by a 50 µl pulse of 1 mM EDTA, all at 50 µl/min. AUC- The sedimentation coefficient for the N9-N1 region (~0.2 mg/ml) was determined by velocity experiments using the Optima XL-A ultracentrifuge (Beckman Instruments). The experiments were performed using double sector cells and a rotor speed of 48000 rpm, taking 150 scans at 1.5 minute intervals at a wavelength of 230 nm and at a temperature of 20 °C. The sedimenting boundaries were analyzed using the program Sedfit version 8.7 (28), and the resulting apparent sedimentation coefficient was corrected for standard conditions using the program Sednterp (29). Hydrodynamic radius (R_h) and frictional ratio (f/f₀), which represents the deviation of the friction of the molecule from a theoretical sphere of the same molecular weight, were also calculated using Sednterp.

**Microfibril purification and composition analysis**- Bovine corneal tissue (approximately 0.2 g wet weight) was finely diced with a razor blade and suspended in 2 ml of buffer (400 mM NaCl, 20 mM Tris-HCl (pH 7.4)) containing protease inhibitors (3 mM N-ethylmaleimide (NEM), 5 mM phenylmethanesulphonyl fluoride (PMSF)) and 0.1 mg/ml chromatographically purified bacterial collagenase (type VII, Sigma, Pool, Dorset, UK). The digestion was left at 4 °C overnight with gentle stirring. The digested tissue was centrifuged at 3000 rpm for 3 minutes, and the supernatant size fractionated on a Sepharose CL-2B column in the same buffer as used for the digest. The V₀ contained collagen VI microfibrils. For MS, the V₀ was concentrated and prepared in a 3 kDa cut-off Nanosep spin column (Pall), pre-conditioned with 25 mM ammonium bicarbonate. Between each step, solutions were removed by centrifugation and discarded. The sample was reduced with 10 mM dithiothreitol for 1 h at 56°C then alkylated with 55 mM iodoacetamide at 20°C for 45 min in the dark. The sample was digested with trypsin (125 ng) at 37°C overnight. Following digestion the solution was retained and the spin column washed with 25 mM ammonium bicarbonate. Between each step, solutions were removed by centrifugation and discarded. The sample was digested with trypsin (125 ng) at 37°C overnight. Following digestion the solution was retained and the spin column washed with 25 mM ammonium bicarbonate. Between each step, solutions were removed by centrifugation and discarded. The sample was digested with trypsin (125 ng) at 37°C overnight. Following digestion the solution was retained and the spin column washed with 25 mM ammonium bicarbonate.

These were combined, dried down and redissolved in 5% (v/v) acetonitrile, 0.1% (v/v) formic acid. Digested samples were analyzed by LC-MS/MS. Data produced were searched using Mascot (Matrix Science) against the IPI Bovine database. Data were validated using Scaffold (Proteome Software).

**Microfibril EM grid preparation and imaging**- Purified microfibrils were adsorbed for 30-60 seconds onto glow-discharged carbon-
coated grids and stained with 2% (w/v) uranyl acetate (pH 4.7). Images were recorded under low dose conditions (<10 e/Å²) on an FEI Tecnai T12 twin EM operating at 120 kV at 52,000x magnification and -0.5 µm defocus on Kodak SO-163 film, and developed for 12 minutes in Kodak D-19 developer at room temperature. Additional images were recorded using low dose conditions on an FEI Polara EM at 200 kV, at nominal magnification 39,000x, giving a calibrated magnification of 50,049x on a Gatan Ultrascan 4000 CCD camera. The defocus range used was -1.5 µm to -3.6 µm.

Generation and refinement of 3D model - The images recorded on the Tecnai T12 twin were digitised at 1500 dpi (3.2 Å pixel resolution) on an Imacon Flextight 848 scanner. 2438 images of the double-bead structure were selected from these data, in 64 x 64 nm boxes. A low-pass Fourier filter was applied at 20 Å resolution. The images were aligned to a common registration using SPIDER (30) procedures. This image set was classified using Imagic5, and the double-bead structure was assessed from the class averages. Then regions containing the half-bead structure were windowed out from the aligned images into 32 x 32 nm boxes. The pairs of images of half-beads together gave a set of ~5000 images. These were refined against an initial 3D model using standard iterative model based refinement in SPIDER (as in (31)). Since the microfibrils were adhered to a carbon support film, images of the microfibrils were assumed to give views perpendicular to the fiber axis. The angular sampling was 5 degrees. As an image quality criterion, a cross-correlation threshold value of 0.35 was applied in the final stages of the refinement. The final map was computed from 1909 images. The resolution is ~20 Å from the Fourier shell correlation (Figure S1).

The initial model was calculated using the Polara data, which included images recorded at 45 degree tilt, as well as untilted images. Initially ~300 particles were selected, centred on the bead structure of the microfibril. Band-pass filtering was used with a high frequency cut-off of 31.3 Å and a low frequency cut-off of 718 Å. These particles were aligned using SPIDER procedures, and classified with Imagic5 (27). Two distinct class averages were chosen and the views were assigned the relative angles of 0 degrees and 45 degrees, perpendicular to the fiber axis, and a low resolution 3D map was computed with two-fold symmetry imposed. The individual images were then refined against this model using the procedures described above, until the parameters converged.

Modeling of the A-domain positions - A homology model of the α2(VI) C2 domain (12) was fitted automatically with UCSF Chimera (32), using the atoms-in-map fitting mode. Thereafter, for the remaining 5, each A-domain was manually placed and the position refined to a local optimum by iterating shift and orientation searches, separately.

RESULTS

Expression of the human α3(VI) N9-N1 region. The non-collagenous N-terminal region of the human α3(VI) chain was expressed in a mammalian expression system and purified as a secreted protein (Figure 1A-C). The first domain, N10, was not included in the recombinant protein as this domain is generally not expressed in human tissues (10,33). Tryptic peptide analysis by MS validated the identity of the purified protein. In order to determine the size and oligomeric status of the N9-N1 region, MALLS was used in conjunction with size exclusion chromatography. The MALLS profile (Figure 1D) showed a single species which had a molecular mass of ~203 kDa ±2%, which is consistent with the predicted mass of a monomer (196,572 Da). This region has only one predicted N-glycosylation site and only a small increase in mobility on SDS-PAGE after PNGaseF digestion (data not shown). Sedimentation velocity AUC was used to confirm that there was a single species with a sedimentation coefficient 7.03 S, f/fo of 1.66 and RH of 6.39 nm (Figure 1E).

Structure of the N9-N1 region

To determine the 3D shape of the protein in solution, SAXS measurements were made at the Daresbury synchrotron radiation source (Figure 2A). The data quality was assessed using Guinier plots, to check for aggregation in the sample (Figure 2B). The radius of gyration (Rg) obtained from the Guinier plot was 6.4 nm. The maximum particle dimension was estimated as 21.4 nm using indirect Fourier transform with GNOM (Figure 2C) (23). Ab initio bead models were generated.
using the program DAMMIN (24). The modelling allowed us to fit the experimental data with discrepancy factor $x$ of $\sim 1.8$. At least 20 separate simulations were completed to determine the common structural features. The mean normalised spatial discrepancy factor between the solutions was 0.9 indicating a unique solution. Ten superimposed $ab$ initio models are shown in Figure 2D, the averaged model had a compact C-shape with dimensions $18 \text{ nm } \times 6 \text{ nm } \times 5.5 \text{ nm}$ (Figure 2E). Since the VWA domain structure is known, we were able to perform rigid body modeling to obtain a more robust model. Nine A-domains were modelled in SASREF (26), using the C2 domain from the $\alpha_2$(VI) chain (12), and fitted the experimental data with a mean discrepancy factor $x$ of 2.1 ($N=14$) (Figure 3A, B). Each simulation produced very reproducible results and the resulting models also had a C-conformation but with a tighter curve than the $ab$ initio model.

The structure of the N9-N1 region was also investigated independently using single particle EM with negative staining (Figure 3C, D). Using reference-free methods, a 3D reconstruction was calculated from 1,776 particles using angular reconstitution. The 3D model generated is similar in overall shape to the models produced from the SAXS analysis (Figure 3E), in particular to the rigid body model, providing further support for a compact C-shaped conformation.

**Metal ion binding of the N9-N1 region.** A recent study highlighted a role for Zn$^{2+}$ in the conformation and ligand binding properties of the A-domain from human matrilin-3 (34), possibly using the MIDAS. Because the N-terminal region of the $\alpha_3$(VI) chain (12) and the $\alpha_2$(VI) chain (12) has a similar dependence. Using the Farfield Analight, we measured the thickness of the N9-N1 region in the absence of metal-ions and in the presence of 500 $\mu$M MgCl$_2$, MnCl$_2$ or ZnCl$_2$. With the addition of 500 $\mu$M MgCl$_2$ or MnCl$_2$ there was a small decrease in the thickness of the protein layer ($\sim 0.2$ nm). With the addition of 500 $\mu$M ZnCl$_2$ there was a larger decrease in thickness of $\sim 0.8$ nm (Figure 4A). The contraction of the protein layer indicates a significant conformational change especially in the presence of ZnCl$_2$. A contraction in the protein layer could be detected at all ZnCl$_2$ concentrations measured (25 - 500 $\mu$M) (Figure 4B).

**Conformational changes in the N9-N1 region in the presence of metal ions.** Next, we used sedimentation velocity AUC to further investigate the conformational change upon metal-ion binding. In the absence of externally applied cations or in the presence of 1mM Mg$^{2+}$ or 1mM Mn$^{2+}$, the protein sediments as a single species with a sedimentation coefficient of 7.03 S (Figure 4C). However, in the presence of 1 mM Zn$^{2+}$, the monomeric species exhibited a higher sedimentation rate (7.93 S) which suggests a more compact conformation, and was accompanied by another species presenting a sedimentation coefficient of 11.9 S (Figure 4C), potentially representing an oligomer.

In order to confirm these findings and to determine the mass of the two species, MALLS was used. In the absence of Zn$^{2+}$ the monomeric protein eluted at a volume of 14 ml and has a molecular weight of 203 kDa $\pm 0.5\%$. However, in the presence of Zn$^{2+}$ two species are seen, one eluting at $\sim 13$ ml and another at $\sim 15$ ml (Figure 4D). The species eluting at 15 ml was confirmed as a monomer (Mwt 197 kDa $\pm 0.3\%$) but the increase in elution volume suggests a more compact conformation in the presence of Zn$^{2+}$, as indicated by AUC. The species eluting at 13 ml had a molecular weight of 435 kDa $\pm 0.8\%$ (Figure 4D) which suggests it is a dimer and presumably correlates with the species sedimenting at 11.9 S in the AUC. Together these data confirm that Zn$^{2+}$ induces the N9-N1 region to undergo a conformational tightening and a degree of higher order molecular assembly.

**3D structure of collagen VI tissue microfibrils.** To determine the organisation of the globular regions in assembled tissue collagen VI microfibrils, we extracted bovine corneal microfibrils using collagenase digestion followed by gel filtration chromatography. Collagen VI is a major component of the ECM of cornea (35) and is resistant to collagenase digestion. After chromatography, MS was used for composition analysis of the Vo fraction. The major component was bovine collagen VI ($\alpha_3$, $\alpha_2$ and $\alpha_1$ chains detected), however there were some other intracellular and extracellular components (Table S3). The only microfibril associated proteins detected were $\beta_{ig}$-h3 and decorin. The $\alpha_6$ chain
was not detected, although this chain is present in mouse cornea (36). Either this chain is not present in the bovine corneal preparation or is below the detection threshold.

Low dose negatively stained EM images were collected and showed that the microfibrils were well preserved (Figure 5A). The characteristic double-beaded structure is clearly visible. The microfibrils are flexible and can bend along the microfibril axis. The beaded regions were extracted from images in 64 x 64 nm boxes, and aligned by single particle methods. In the average of the aligned bead region and in class averages (Figure 6A) a distinct intrabead gap is resolved, separating the double-bead into two halves which we have named half-beads.

A level of heterogeneity was observed, in the class averages of the double-bead images, as in-plane twisting between the half-beads. The individual half-bead structures appear ordered, and the largest movement occurs at the junction between half-beads (movie S4). Since the collagen VI tetramer is two-fold symmetric, a dimer of dimers, each half-bead should be an equivalent structure (see Figure 7). Therefore, for further analysis, the images of the half-beads were extracted and aligned (Figure 6B shows selected class averages from the single particle refinement).

A 3D map of the half-bead was computed and shows that this structure is composed of 4 distinct layers (Figure 6C). The half-bead structure measures roughly 20 nm in length. There is a well defined, compact and distinctly dome-shaped head region, followed by an intermediate layer, which leads into an extended tail region with radiating spoke-like features. The head region corresponds to layer I in Figure 6C. The intermediate region forms layer II, while the tail region is comprised of layers III and IV.

The head and intermediate region are compact, while the tail region is more variable. The head layer is slightly oval, but approximately round with diameter of 9.6 nm, while the intermediate region is more rectangular in shape and measures approximately 6.5 nm x 8.6 nm (Figure 8A). The diameter of the tail region is in the range 6.0 nm to 15.3 nm. The dome-shaped head region is hollow (Figure 8B) and is fairly homogeneous, as can be seen in the averages in Figure 6A. The head region would correspond to the C-terminal region and the more variable tail region to the N-terminal region, in current assembly models. The structure and class averages indicate a well defined and organized arrangement, as would be expected for the complex hierarchical assembly.

**DISCUSSION**

Using a multidisciplinary approach, we have combined solution biophysical approaches with EM image analysis to describe the 3D nanostructure of nine contiguous A-domains from collagen VI. All these techniques describe a compact arrangement of domains and converge on a single shape. This region has a compact "C"-shape with dimensions 18 nm x 6 nm x 5.5 nm (Figure 3). A previous study analyzed an N9-N2 construct using rotary shadowing and negative staining EM. This study suggested two packing models compatible with their 2D image data, a zig-zag conformation and a folded conformation (16). Our 3D reconstruction indicates that the actual shape bears some similarities to both these models. Our earlier electron tomography study showed that a hexagonal arrangement of globular domains was occasionally present on the surface of the microfibril bead (6). This particular arrangement of domains can be recreated by projecting the N9-N1 model in certain orientations. This arrangement is only seen in a small number of projections, which would explain its infrequent occurrence in the tomography reconstructions.

In the absence of externally added metal-ions the N9-N1 region is monomeric, which suggests this region does not drive oligomerization in the absence of other factors. We investigated the role of cation binding on the shape and oligomeric state of the N9-N1 region as the N4 domain contains one complete MIDAS sequence. Mn$^{2+}$ and Mg$^{2+}$ ions had a small effect, measured by Farfield and quartz crystal microbalance (Figure S6), which however did not result in a conformational change as analysed by AUC. However, Zn$^{2+}$ ions induced changes to the N9-N1 region which were detected by Farfield and AUC. In the presence of Zn$^{2+}$ ions, the N9-N1 region becomes more compact which is consistent with the recent findings for matrilin-3, a VWA domain containing cartilage protein (34). In addition, a proportion of the collagen VI A-domains dimerise
in the presence of Zn$^{2+}$ ions. The intermolecular interactions between $\alpha$3 N-terminal domains could reflect associations between dimers in microfibril formation (see Figure 7 for assembly model); although there is no evidence that this is a metal-ion dependent process. Zn$^{2+}$ has an important role in maintaining ocular function and in cartilage formation (37,38), although we do not currently know whether Zn$^{2+}$ binding has a physiological role in collagen VI assembly or function. Zn$^{2+}$ ions are unlikely to be liganded by the MIDAS and typically interact with cysteine residues. There are 4 cysteine residues in the N9-N1 region which may play a role in Zn$^{2+}$ binding.

We used collagen VI microfibrils isolated from bovine cornea to determine their 3D structure. Initially we used the entire double-bead region for our analysis but it became apparent that there was conformational flexibility along the microfibril axis and an element of twisting within each double-bead. We then decided to use each half-bead as our basic unit. The half-bead is organized into four distinct layers, which are oriented perpendicular to the microfibril axis (Figure 6C). Measurements of the half-bead are compatible with the overlapping interaction of tetramers to form the microfibril (see Figure 7 for explanation). This is the model most recently described in (6) and places the C-terminal domains in the head region and the N-terminal domains of the adjoining tetramer at the tail.

The N-terminus of the $\alpha$3 chain contains up to 10 VWA domains, making the N-terminal region more variable than the C-terminus. The C-terminus of the $\alpha$3 chain has 3 small additional domains, which although essential for microfibril formation (13) are thought to be cleaved after microfibril formation (39-41). To ascertain which domains are present in these microfibrils, the peptides detected by MS were analyzed (Figure S7). In the $\alpha$3 chain, no peptides were detected from the N10 domain, which is not unexpected considering that this domain is almost always spliced out (33,42). Only one peptide was detected from each of the N9 and N7 domains which are also subject to variable splicing. However for each of the other N-terminal domains at least 3 peptides were detected. No peptides were detected after residue 2496 in the C1 domain of the $\alpha$3 chain which is consistent with previous tissue extraction studies (40,41) and suggests that this region may be cleaved. However peptides were detected in the C2 domain of both the $\alpha$1 and $\alpha$2 chains suggesting that processing only affects the $\alpha$3 chain. The molecular weight estimated from SDS-PAGE is consistent with an $\alpha$3 chain composed of N8-N6 – C1 (see Figure 5B).

The placement of the C1 domains in the head of the half-bead structure has been modeled using a homology modeled collagen VI A-domain (12). Although six A-domains fit well into the head, it is not 6-fold symmetric (Figure 8C). The head is slightly oval shaped and consistent with two triple-helical monomers. The C2 domains are probably placed in the next layer of the structure, but this region is more poorly defined and contains less density. The intermediate layer is rectangular in shape, rather than round or oval. This shape is consistent with four C2 domains from the $\alpha$1 and $\alpha$2 chains in this layer, assuming cleavage of the $\alpha$3 chain C2-C5 domains. This region is possibly where N- and C-terminal domains interact and the layers below this (in Figure 8) are likely to be composed of N-terminal domains.

The collagenuous region runs into the head region of the half-bead. The hollow head indicates that the collagenuous region must take an external path around the bead, and that the globular domains are not organized around a central collagenuous core. A strong feature present between the two half-beads, highlighted in Figure 6A, is likely to be the collagenuous region. A protrusion from the outer parts of the tails (Figure 8A and B) could be due to the collagenuous region running past the half-bead. It is not visible to a greater extent due to the structural heterogeneity in the tail region.

The tail region is clearly more variable and resembles the structure of the recombinant N-terminal region, as shown in Figure 3, in character. As in the structure of the isolated N-terminal region, the tails appear as a linked series of blobs, which are compatible with the size of the modeled N-terminal domains. The N-terminal A-domains of the $\alpha$3 chain are required and a minimal N1-N5 appears necessary to support microfibril formation (15). In addition, the recently discovered $\alpha$4, $\alpha$5 and $\alpha$6 chains all have an extended N-terminus and probably assemble with the shorter $\alpha$1 and $\alpha$2 chains (1,2). Together these data highlight the importance of the extended N-terminal region for microfibril assembly. The N-terminal regions
protrude away from the beaded region and it is tempting to indicate a role for these domains in mediating supramicrofibrillar interactions or with other matrix proteins.

REFERENCES

1. Gara, S. K., Grumati, P., Urciuolo, A., Bonaldo, P., Kobbe, B., Koch, M., Paulsson, M., and Wagener, R. (2008) *J Biol Chem* **283**, 10658-10670
2. Fitzgerald, J., Rich, C., Zhou, F. H., and Hansen, U. (2008) *J Biol Chem* **283**, 20170-20180
3. Lampe, A. K., and Bushby, K. M. (2005) *J Med Genet* **42**, 673-685
4. Baker, N. L., Morgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F., and Lamande, S. R. (2005) *Hum Mol Genet* **14**, 279-293
5. Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E., and Engel, J. (1983) *Biochem J* **211**, 303-311
6. Baldock, C., Sherratt, M. J., Shuttleworth, C. A., and Kielty, C. M. (2003) *J Mol Biol* **330**, 297-307
7. Engel, J., Furthmayr, H., Odermatt, E., von der Mark, H., Aumailley, M., Fleischmajer, R., and Timpl, R. (1985) *Ann NY Acad Sci* **460**, 25-37
8. Chu, M. L., Pan, T. C., Conway, D., Kuo, H. J., Glanville, R. W., Timpl, R., Mann, K., and Deutzmann, R. (1989) *Embo J* **8**, 1939-1946
9. Chu, M. L., Pan, T. C., Conway, D., Saitta, B., Stokes, D., Kuo, H. J., Glanville, R. W., Timpl, R., Mann, K., and Deutzmann, R. (1990) *Ann NY Acad Sci* **580**, 55-63
10. Zanussi, S., Doliana, R., Segat, D., Bonaldo, P., and Colombatti, A. (1992) *J Biol Chem* **267**, 24082-24089
11. Dickeson, S. K., and Santoro, S. A. (1998) *Cell Mol Life Sci* **54**, 556-566
12. Ball, S. G., Baldock, C., Kielty, C. M., and Shuttleworth, C. A. (2001) *J Mol Biol* **276**, 7422-7430
13. Lamande, S. R., Morgelin, M., Adams, N. E., Selan, C., and Allen, J. M. (2006) *J Biol Chem* **281**, 16607-16614
14. Tooley, L. D., Zamurs, L. K., Beecher, N., Baker, N. L., Peat, R. A., Adams, N. E., Bateman, J. F., North, K. N., Baldock, C., and Lamande, S. R. (2010) *J Biol Chem* **285**, 33567-33576
15. Fitzgerald, J., Morgelin, M., Selan, C., Wiberg, C., Keene, D. R., Lamande, S. R., and Bateman, J. F. (2001) *J Biol Chem* **276**, 187-193
16. Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J., and Chu, M. L. (1992) *Embo J* **11**, 4281-4290
17. Hanssen, E., Reinboth, B., and Gibson, M. A. (2003) *J Biol Chem* **278**, 24334-24341
18. Finnis, M. L., and Gibson, M. A. (1997) *J Biol Chem* **272**, 22817-22823
19. Wiberg, C., Hedbom, E., Khairullina, A., Lamande, S. R., Oldberg, A., Timpl, R., Morgelin, M., and Heinegard, D. (2001) *J Biol Chem* **276**, 18947-18952
20. Wiberg, C., Klatt, A. R., Wagener, R., Paulsson, M., Bateman, J. F., Heinegard, D., and Morgelin, M. (2003) *J Biol Chem* **278**, 37698-37704
21. Poschl, E., Fox, J. W., Block, D., Mayer, U., and Timpl, R. (1994) *EMBO J* **13**, 3741-3747
22. Berry, R., Jowitt, T. A., Ferrand, J., Roessle, M., Grossmann, J. G., Canty-Laird, E. G., Kammerer, R. A., Kadler, K. E., and Baldock, C. (2009) *Proc Natl Acad Sci USA* **106**, 8561-8566
23. Semenyuk, A. V., Svergun, D.V. (1991) *Journal of Applied Crystallography* **24**, 537-540
24. Svergun, D. I. (1999) *Biophys J* **76**, 2879-2886
25. Volkov, V. V., Svergun, D.I. (2003) *Journal of Applied Crystallography* **36**, 860-864
26. Petoukhov, M. V., and Svergun, D. I. (2005) *Biophys J* **89**, 1237-1250
27. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996) *J Struct Biol* **116**, 17-24
28. Schuck, P. (2000) *Biophys J* **78**, 1606-1619
29. Philo, J. S. (1997) *Biophys J* **72**, 435-444
30. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladadj, M., and Leith, A. (1996) *J Struct Biol* **116**, 190-199
31. Roseman, A. M., Chen, S., White, H., Braig, K., and Saibil, H. R. (1996) *Cell* **87**, 241-251
32. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**, 1605-1612
33. Dziadek, M., Kazenwadel, J. S., Hendrey, J. A., Pan, T. C., Zhang, R. Z., and Chu, M. L. (2002) *Matrix Biol* **21**, 227-241
34. Fresquet, M., Jowitt, T. A., Ylostalo, J., Coffey, P., Meadows, R. S., Ala-Kokko, L., Thornton, D. J., and Briggs, M. D. (2007) *J Biol Chem* **282**, 34634-34643
35. Zimmermann, D. R., Trueb, B., Winterhalter, K. H., Witmer, R., and Fischer, R. W. (1986) *FEBS Lett* **197**, 55-58
36. Gara, S. K., Grumati, P., Squarzoni, S., Sabatelli, P., Urciuolo, A., Bonaldo, P., Paulsson, M., and Wagener, R. (2011) *Matrix Biol* **30**, 248-257
37. Grahn, B. H., Paterson, P. G., Gottschall-Pass, K. T., and Zhang, Z. (2001) *J Am Coll Nutr* **20**, 106-118
38. Rosenberg, K., Olsson, H., Morgelin, M., and Heinegard, D. (1998) *J Biol Chem* **273**, 20397-20403
39. Aigner, T., Hambach, L., Soder, S., Schlotzer-Schrehardt, U., and Poschl, E. (2002) *Biochem Biophys Res Commun* **290**, 743-748
40. Mayer, U., Poschl, E., Nischt, R., Specks, U., Pan, T. C., Chu, M. L., and Timpl, R. (1994) *Eur J Biochem* **225**, 573-580
41. Trueb, B., and Winterhalter, K. H. (1986) *EMBO J* **5**, 2815-2819
42. Stokes, D. G., Saitta, B., Timpl, R., and Chu, M. L. (1991) *J Biol Chem* **266**, 8626-8633
43. Abedinia, M., Pain, T., Algar, E. M., and Holmes, R. S. (1990) *Exp Eye Res* **51**, 419-426

**FOOTNOTES**

We would like to thank Mrs Elizabeth Lord for excellent technical assistance. We would like to acknowledge Dr David Knight, Marge Howard and Emma Keевill in the Biomolecular analysis facility, Adam Huffman in the Bioinformatics facility and the staff in the EM facility, in particular Drs David Holmes and Tobias Starborg (University of Manchester). We would like to thank Professor Mon Li Chu, Thomas Jefferson University, for collagen VI DNA. NB was supported by a BBSRC grant (Ref: D008662) to CB, CMK and CAS, and RB by a BBSRC studentship (Ref: BBSSC200513125) to CB. The FEI Polara was funded by the Wellcome Trust (grant ref: WT081406MA). We would also like to thank Daresbury SRS for beamtime and Dr J.G. Grossmann for assistance during data collection.

**ABBREVIATIONS**

The abbreviations used are: SAXS, small angle X-ray scattering; AUC, analytical ultracentrifugation; MALLS, multilateral laser light scattering; MIDAS, metal-ion dependent adhesion site; UCMD, Ullrich congenital muscular dystrophy; ECM, extracellular matrix; VWA, von Willebrand Factor type A domains, βig-h3, transforming growth factor-β induced gene-h3.
**FIGURE LEGENDS**

**Figure 1** Expression and characterization of the human α3(VI) N9-N1 region. (A) Schematic diagram of the α3(VI) N9-N1 construct. Squares represent the VWA domains. The construct has a C-terminal 10x histidine tag following a thrombin cleavage sequence. (B) Coomassie stained SDS-PAGE of the purified N9-N1 region under reducing conditions. Lane 1 is the molecular weight marker and lane 2 shows the N9-N1 region running as a single band of ~190 kDa after elution from the Ni-NTA column but prior to gel filtration. (C) Size exclusion chromatography of the N9-N1 region. The graph shows the absorbance at 280 nm wavelength with volume. The major species elutes from the size exclusion column at 11.5 ml. (D) Multiangle laser light scattering of the N9-N1 region. The graph shows the differential refractive index and molecular mass with volume. The molecular mass of 203,000 ± 4,060 Da (experimental errors from polydispersity) is approximately that expected for a monomer (195,572 Da). (E) C(s) analysis of the N9-N1 region as derived from sedimentation velocity AUC.

**Figure 2** Solution small angle X-ray scattering data for the α3(VI) N9-N1 region. (A) The experimental SAXS data for the N9-N1 region are plotted as a function of q. (B) The low-angle region of the X-ray scattering data is show in the form of a Guinier plot, which is linear for values q ≤ 1/Rg (black squares). (C) The distance distribution function is shown. (D) Shapes were simulated *ab initio* by the programs DAMMIN. An example of 10 independent DAMMIN simulations are shown superimposed to highlight the uniqueness of the solution. (E) These were used to calculate an average "most probable" shape which is represented as a solid model.

**Figure 3** Nanostructure of the α3(VI) N9-N1 region. (A) Rigid body modeling of nine A-domains was performed to the experimental SAXS data using the program SASREF. A representative model is shown in two orthogonal views (i and ii) with the domains shown in a cartoon representation and colored blue at the N-terminal domain through to red at the C-terminal domain. (B) Graph of the fit between the experimental SAXS data and the theoretical X-ray scattering of the rigid body model shown in (A). (C) Low dose negatively stained EM images were recorded on a Tecnai T12 twin at 120 kV using a CCD detector. Individual N9-N1 particles were windowed into 26 x 26 nm boxes. Selected representative class averages from the single particle image processing are shown. (D) Images of the 3D reconstruction represented as a solid surface are shown. (E) An overlay of the EM 3D reconstruction is shown as a mesh, with a rigid body model from the SAXS analysis superimposed.

**Figure 4** Interaction with metal-ions by the α3(VI) N9-N1 region. (A) The change in thickness of the N9-N1 region was measured in the presence of either 500 µM Mg^{2+}, Mn^{2+}, or Zn^{2+} using the Farfield Analight. (B) Table showing the decrease in thickness of the N9-N1 layer with increasing concentrations of Zn^{2+}. These data are also expressed as a percentage change from the starting thickness. (C) C(s) analysis of the N9-N1 region in the absence and presence of 1 mM concentrations of either Mg^{2+}, Mn^{2+}, and Zn^{2+} as derived from sedimentation velocity AUC. (D) Multiangle laser light scattering of the N9-N1 region in the presence and absence of Zn^{2+}. The graph shows the differential refractive index with volume. In the absence of Zn^{2+} the major species elutes from the size exclusion column at 14 ml but in the presence of Zn^{2+} there are two species, a minor one at 13.25 ml and another at 14.75 ml. The molecular mass was plotted for each species.

**Figure 5** Extraction and negative staining EM of bovine corneal collagen VI microfibrils. (A) Electron micrograph image of a field of collagen VI microfibrils stained with uranyl acetate, taken on the FEI Polara at 200 kV, magnification ~50,000 x. The image was recorded on a Gatan Ultrascan 4000 CCD. No image enhancement or other filters have been applied. The arrow indicates a distinctive image that resembles one of the class averages shown in Figure 6. A filamentous line of density appears to take an external route around the particle. Scale bar = 100 nm. (B) Coomassie stained SDS-PAGE of the extracted collagen VI microfibrils under reducing conditions. Lane 1 is the molecular weight marker and
lane 2 shows the three bands seen in the collagen VI preparation. These were identified by MS as the collagen VI α1, α2 and α3 chains and aldehyde dehydrogenase, the major soluble protein in cornea (43). The most abundant α3 chain in these preparations is likely to be N8–N6 – C1.

Figure 6 3D reconstruction of the half-bead region from collagen VI microfibrils. (A) The average of all collagen VI “double-beads”, followed by 23 representative class averages. 23 of 50 class averages are shown, representing 1055 out of 2438 images. The box size is 64 x 64 nm. The double-bead has approximate dimensions of ~15 nm x 45 nm. The globular intrabead feature is indicated by the arrow in the first frame, this can also be seen in some of the class averages. Conformation variability is apparent in these images; different views of the half-beads and angles between the individual half-beads can be seen. (B) Panel of 9 representative class averages of images from the single particle refinement, the box size is 32 x 32 nm. Images were recorded on FEI Tecnai T12 twin at 120 kV. Some of the views can be recognized in Figure 5. A full set of class averages and corresponding projections of the structure is shown in Supplementary Figure S5. (C) Radial distribution of density from a cylindrically averaged structure. The collagen VI half-bead has a layered structural organization. A slice through the center of the radially averaged 3D density (left) when projected to a 1D trace (right) displays a distinctly banded structure. They have been assigned and labeled from the top as bands I (head), II (intermediate), III and IV (tail regions). These bands measure 5.4, 3.8, 4.8 and 7.0 nm respectively. The scale on the right is in Angstroms. Produced in SPIDER WEB.

Figure 7 Schematic diagrams showing the assembly of collagen VI. (A) Domain organization of the three alpha chains. Blue and red circles represent N- and C-terminal VWA domains, respectively. (B) The α1, α2 and α3 chains assemble to form a triple-helical monomer. (C) Dimers are formed by the anti-parallel overlapping association of two monomers. The overlapping collagenous region is represented as a solid line whereas the non-overlapping collagenous regions are shown as dotted lines. (D) Tetramers are formed by the parallel association of two dimers. (E) Microfibrils are formed by end-to-end interaction of tetramers. Globular regions from one tetramer are labeled N_a (blue), C_a (red), and from overlapping tetramers N_b, N_c (grey), C_b, C_c (pink). The overlapping collagenous region is represented as a solid line whereas the non-overlapping collagenous regions are shown as dotted lines. The beaded region boxed out for EM analysis has been indicated. Modified from (6).

Figure 8 Modeling VWA-domains into the half-bead structure. (A) External views of the 3D map are shown in two different orientations ((i) and (ii)). The surface is colored along the microfibril axis, red (head) to blue (tail), corresponding to assignment of the C- and N-termini of collagen VI (see text for explanation). (B) Cut open view of the 3D map colored by relative density of the 3D map. The color key indicates increasing density from blue to red. Maps are rendered with UCSF Chimera. (C) 3D map of the half-bead with 6 VWA-domains fitted to the head part of the structure. (i) Six VWA-domains, representing the C1 domains from two triple-helical monomers, were docked into the 3D map of the half-bead. (ii and iii) Cut open views of the 3D map showing 3 A-domains fitted in each half. Docking was performed with UCSF Chimera and the map was also rendered with UCSF Chimera. B and C show the hollow center of the head layer.
Figure 2

(A) Plot of Log I vs. q (Å⁻¹) with data points.

(B) Plot of Log I vs. q² (Å⁻²) showing a linear relationship.

(C) Graph of P(r) vs. Dmax (Å) with a peak at around 150 Å.

(D) Three-dimensional representation of a molecular structure.

(E) Another three-dimensional representation of a different molecular structure.
Figure 3

A(i) Experimental data
B Experimental data
C Sasref fit
D
E

Log I

q (Å⁻¹)

1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
0.00
0.05
0.10
0.15
0.20
0.25

0

18 nm

18 nm

14

by guest on March 24, 2020 http://www.jbc.org/Downloaded from
Figure 4

A. Graph showing thickness decrease (nm) over time (seconds) for 500 uM Zn, 500 uM Mg, and 500 uM Mn.

B. Table showing thickness decrease (nm) and % thickness change for different [Zn²⁺] concentrations:

| [Zn²⁺] (µM) | Thickness decrease (nm) | % Thickness change |
|-------------|-------------------------|--------------------|
| 25          | 0.272                   | 4.08               |
| 50          | 0.328                   | 4.92               |
| 100         | 0.438                   | 6.56               |
| 200         | 0.502                   | 7.52               |
| 300         | 0.568                   | 8.51               |
| 400         | 0.674                   | 10.09              |
| 500         | 0.775                   | 11.61              |

C. Sedimentation coefficient graph for WT, WT + Mg, WT + Mn, and WT + Zn.

D. Graph showing molar mass (g/mol) and refractive index (RI) for different conditions.
Figure 5

A

B

\[ \alpha_3 (VI) \]

\[ \alpha_1 + \alpha_2 (VI) \]

aldehyde dehydrogenase

Collagenous region

\[ \alpha_3 (VI) \]

N-ter \( N_8 \) \( N_6 \) \( N_5 \) \( N_4 \) \( N_3 \) \( N_2 \) \( N_1 \) C-ter
Figure 7

A

\[ \alpha_1 \mid N_1 \mid C_1 \mid C_2 \]

\[ \alpha_2 \mid N_1 \mid C_1 \mid C_2 \]

\[ \alpha_3 \mid N_{10} \mid N_1 \mid C_1 \mid C_2 \mid C_3 \mid C_5 \]

Collagenous region

Proline-rich repeat

Type III Fn repeat

Kunitz domain

B

Triple-helical monomer

\[ N \]

C

Dimer

\[ \text{overlapped collagenous region} \]

\[ \text{non-overlapped collagenous region} \]

D

Tetramer

E

Microfibril

\[ N_a \]

\[ C_5 \]

\[ C_3 \]

\[ N_b \]

\[ C_6 \]

\[ N_c \]

\[ C_7 \]

~47 nm

Half-bead

Double-bead
Collagen VI: Conformation of A-domain arrays and microfibril architecture
Nicola Beecher, Alan M. Roseman, Thomas A. Jowitt, Richard Berry, Helen Troilo, Richard A. Kammerer, C. Adrian Shuttleworth, Cay M. Kielty and Clair Baldock

J. Biol. Chem. Published online September 9, 2011

Access the most updated version of this article at doi: 10.1074/jbc.M111.265595

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/09/09/M111.265595.DC1