⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 46, pp. 40266–40275, November 18, 2011

Collagen VI, Conformation of A-domain Arrays and Microfibril Architecture[S]

Received for publication, May 27, 2011, and in revised form, July 11, 2011 Published, JBC Papers in Press, September 9, 2011, DOI 10.1074/jbc.M111.265595

Nicola Beecher[‡1,2], Alan M. Roseman[§2], Thomas A. Jowitt[‡], Richard Berry[‡3], Helen Troilo[‡], Richard A. Kammerer[‡4], C. Adrian Shuttleworth[‡], Cay M. Kielty[‡], and Clair Baldock[‡5]

From the [‡]Wellcome Trust Centre for Cell-Matrix Research and [§]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

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 10 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. Furthermore, using biophysical approaches, we demonstrate that the N-terminal region undergoes a conformational change and a proportion forms dimers in the presence of Zn^{2+}. This is the first indication that divalent cations interact with collagen VI A-domains. A three-dimensional 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)[6] 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 Ullrich congenital muscular dystrophy and Bethlem myopathy (for review, see Ref. 3). Bethlem myopathy is at the mild end of the clinical spectrum and is associated with reduced collagen VI levels, whereas Ullrich congenital muscular dystrophy 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 and 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-terminal and two C-terminal domains, the α3(VI) chain is much larger with a maximum of 10 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 sulfate, hyaluronan (16), the transforming growth factor-β induced gene-h3 (βig-h3) (17), microfibril-associated glycoprotein-1 (18), 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

⌘ *Author's Choice*—Final version full access.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1, Figs. S1–S5, and Movie S1.

[1] Supported by Biotechnology and Biological Sciences Research Council Grant D008662 (to C. B., C. M. K., and C. A. S.)

[2] Both authors contributed equally to this work.

[3] Supported by Biotechnology and Biological Sciences Research Council Studentship BBSSC200513125 (to C. B.).

[4] Present address: Laboratory of Biomolecular Research, OFLC/106, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland.

[5] To whom correspondence should be addressed: 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.

[6] The abbreviations used are: ECM, extracellular matrix; SAXS, small angle x-ray scattering; AUC, analytical ultracentrifugation; MALLS, multiangle laser light scattering; MIDAS, metal ion-dependent adhesion site; VWA, von Willebrand Factor type A domains.

the role of these additional domains 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 utilized 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 three-dimensional reconstruction. The three-dimensional 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 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 Ref. 16 (gift from Professor M. L. Chu). A His_{10} 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 (1 unit/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 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 LC-MS/MS using a NanoAcquity LC (Waters) coupled to a LTQ Velos (Thermo Fisher Scientific).

MALLS Analysis—Samples (0.5 ml at ~0.5 mg/ml) were loaded onto a Superdex200 10/300GL column running at a flow rate of 0.5 ml/min in 50 mM 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 analyzed using Astra (version 5.3.2).

Solution SAXS—Data on the N9-N1 region at ~5 mg/ml were collected at station 2.1 of the Synchrotron Radiation Source at Daresbury Laboratory. The sample was maintained at 10 °C during data collection. Images were collected in multiple 60-s 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, R_g and one-dimensional intraparticle 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 software package (25). Rigid body modeling to the experimental scattering data were 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 ($<10\ e^-/Å^2$) on a 2048 × 2048 pixel CCD camera at 52,000× magnification between −0.5 and −1.0 μm defocus. Images were converted to Imagic5 format (Image Science); the Imagic5 suite of programs was used for particle picking and image processing (27). The total number of particles in the data set was 1776. 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 data set during iterative rounds of translational and rotational alignment. Symmetry was not applied at any stage of the image analysis to the data set. Euler angles were assigned to class-sum images, enabling calculation of an initial three-dimensional reconstruction, which was then subjected to multiple rounds of iterative refinement.

FarField Methods—Dual polarization interferometry was performed using an Analight four-dimensional 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 $MgCl_2$, $MnCl_2$, and $ZnCl_2$ and then $ZnCl_2$ 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 48,000 rpm, taking 150 scans at 1.5-min 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). The hydrodynamic radius (R_H) and frictional ratio (f/f_o), 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 (~0.2 g of 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

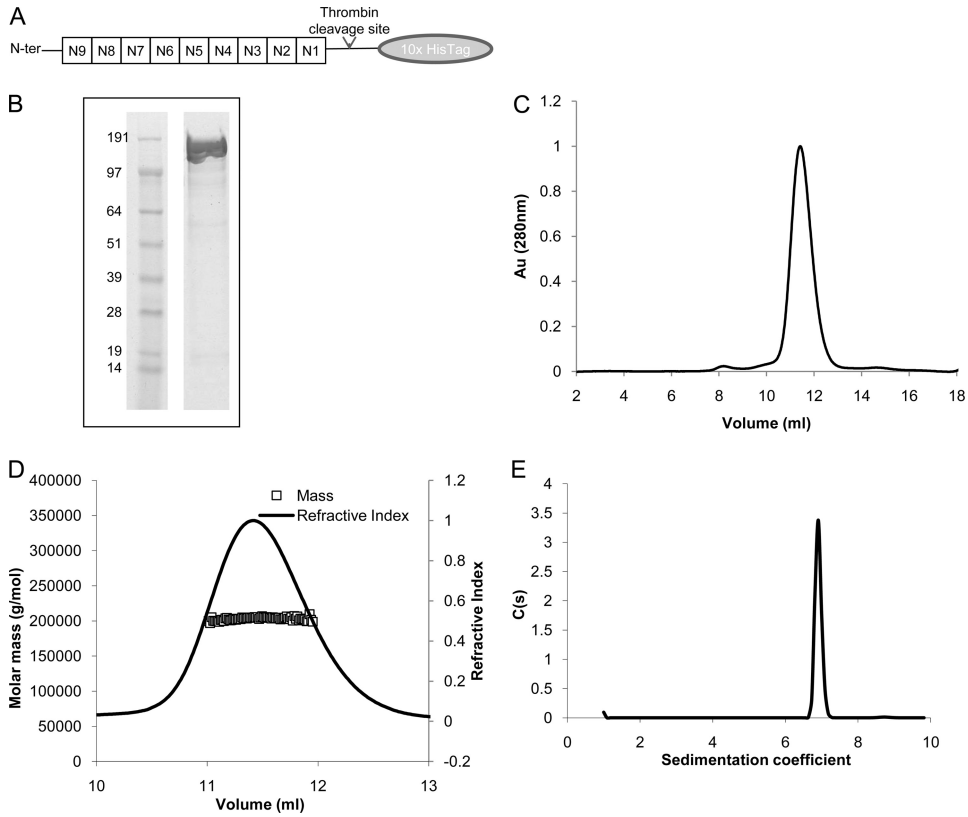


FIGURE 1. **Expression and characterization of the human α3(VI) N9-N1 region.** *A*, schematic diagram of the α3(VI) N9-N1 construct. *Squares* represent the VWA domains. The construct has a C-terminal His_{10} tag following a thrombin cleavage sequence. *B*, Coomassie stained SDS-PAGE of the purified 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 nickel-nitrilotriacetic acid 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. *Au*, absorbance units. *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 ± 4060 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.

N-ethylmaleimide, 5 mM PMSF) and 0.1 mg/ml chromatographically purified bacterial collagenase (type VII, Sigma). The digestion was left at 4 °C overnight with gentle stirring. The digested tissue was centrifuged at 3000 rpm for 3 min, and the supernatant size fractionated on a Sepharose CL-2B column in the same buffer as used for the digest. The V_o contained collagen VI microfibrils. For MS, the V_o 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 and 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 was washed with 25 mM ammonium bicarbonate. These were combined, dried, 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 absorbed for 30–60 s onto glow-discharged carbon-coated grids and stained with 2% (w/v) uranyl acetate (pH 4.7). Images were recorded under low dose conditions ($<10\ e^-/Å^2$) on an FEI Tecnai T12 twin EM operating at 120 kV at 52,000× magnification and −0.5 μm defocus on Kodak SO-163 film and developed for 12 min 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 a nominal magnification of 39,000×, giving a calibrated magnification of 50,049× on a Gatan Ultrascan 4000 CCD camera. The defocus range used was −1.5 μm to −3.6 μm.

Generation and Refinement of Three-dimensional Model—The images recorded on the Tecnai T12 twin were digitized 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 × 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 × 32 nm boxes. The pairs of images of half-beads together gave a set of ~5000 images. These were refined against an initial three-dimensional model using standard iterative model based refinement in SPIDER (as in Ref. 31). Because 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 °. As an

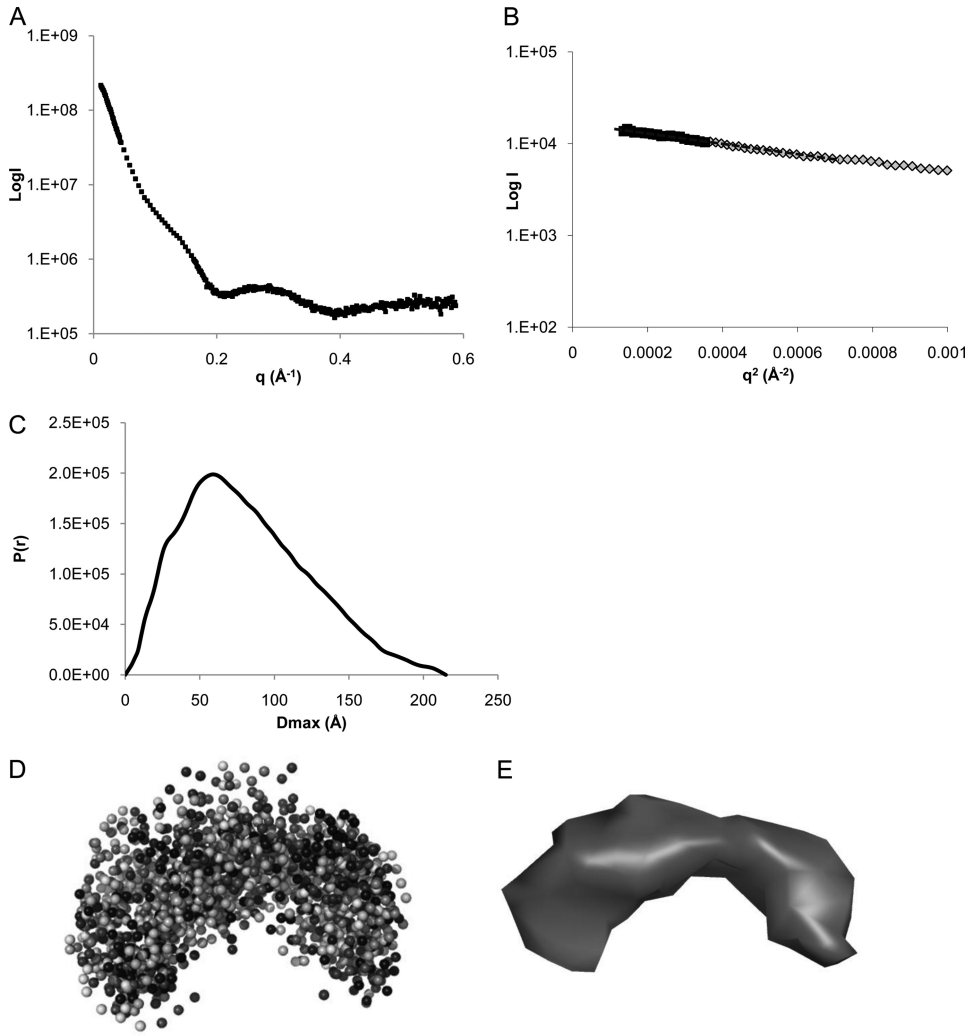


FIGURE 2. **Solution small angle x-ray scattering data for the α3(VI) N9-N1 region.** *A*, the experimental SAXS data for the N9-N1 region are plotted as a function of *q*. *B*, the low-angle region of the x-ray scattering data is show in the form of a Guinier plot, which is linear for values $q \leq 1/R_g$ (*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.

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 (supplemental Fig. S1).

The initial model was calculated using the Polara data, which included images recorded at a 45 ° tilt, as well as untilted images. Initially, ~300 particles were selected, centered 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 ° and 45 °, perpendicular to the fiber axis, and a low resolution three-dimensional map was computed with 2-fold symmetry imposed. The individual images were then refined against this model using the procedures described above, until the parameters converged.

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

RESULTS

Expression of 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 (Fig. 1, *A–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. To determine the size and oligomeric status of the N9-N1 region, MALLS was used in conjunction with size exclusion chromatography. The MALLS profile (Fig. 1*D*) 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 PNGase F

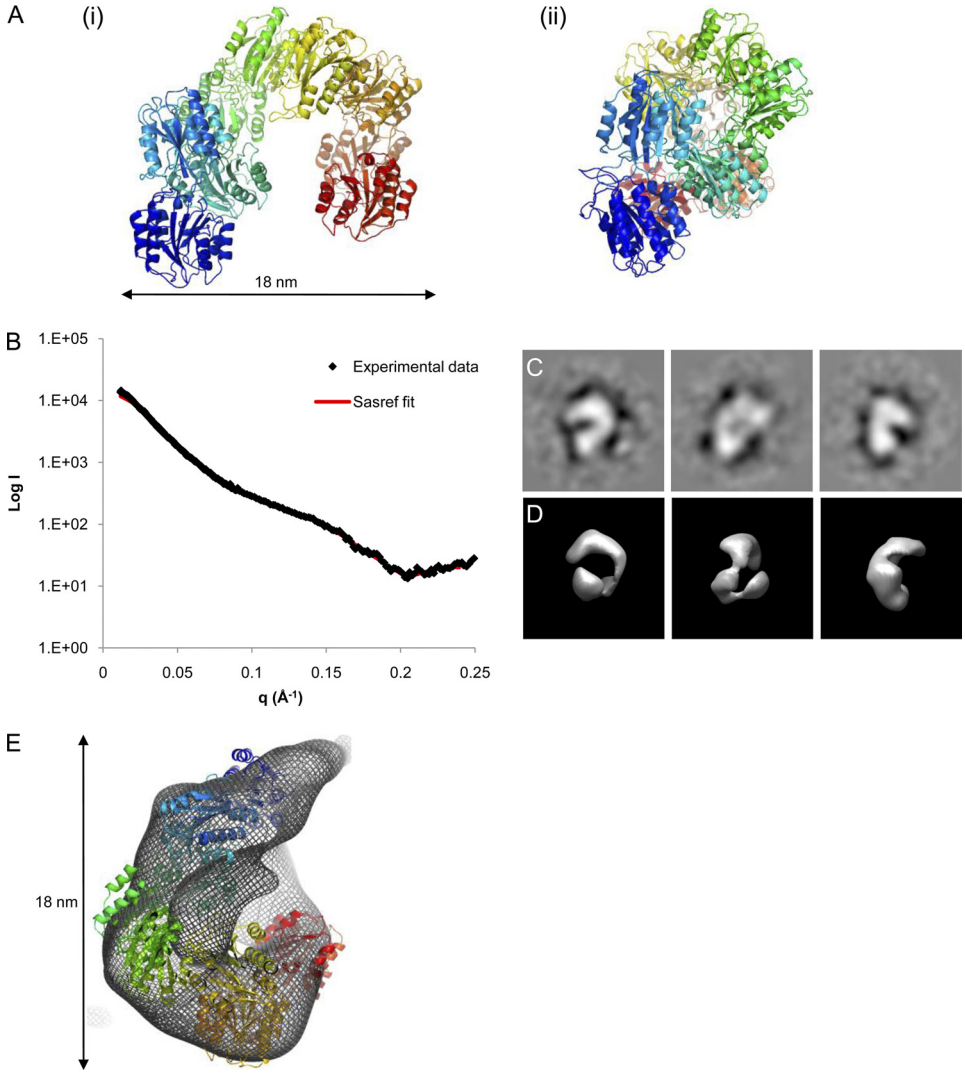
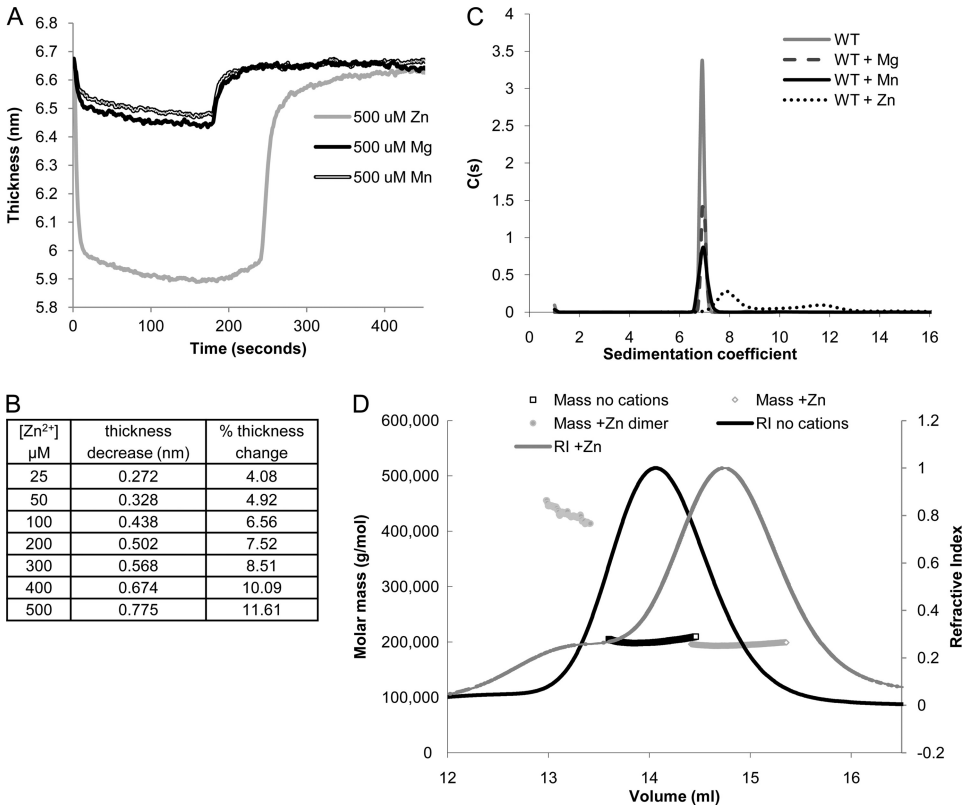


FIGURE 3. **Nanostructure of the α3(VI) N9-N1 region.** *A*, rigid body modeling of nine A-domains was performed to the experimental SAXS data using the program SASREF. A representative model is shown in two orthogonal views (*panels i* and *ii*) with the domains shown in a schematic representation and 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 × 26 nm boxes. Selected representative class averages from the single particle image processing are shown. *D*, images of the three-dimensional reconstruction represented as a solid surface are shown. *E*, an overlay of the EM three-dimensional reconstruction is shown as a *mesh*, with a rigid body model from the SAXS analysis superimposed.

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/f_o of 1.66 and R_H of 6.39 nm (Fig. 1*E*).

Structure of N9-N1 Region—To determine the three-dimensional shape of the protein in solution, SAXS measurements were made at the Daresbury synchrotron radiation source (Fig. 2*A*). The data quality was assessed using Guinier plots, to check for aggregation in the sample (Fig. 2*B*). The radius of gyration (R_g) 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 (Fig. 2*C*) (23). *Ab initio* bead models were generated using the program DAMMIN (24). The modeling allowed us to fit the experimental data with discrepancy factor x of ~1.8. At least 20 separate simulations were completed to determine the common structural features. The mean normalized spatial discrepancy factor between the solutions was 0.9 indicating a unique solution. Ten superimposed *ab initio* models are shown in Fig. 2*D*, the averaged model had a compact C-shape with dimensions 18 × 6 × 5.5 nm (Fig. 2*E*). Because the VWA domain structure is known, we were able to perform rigid body modeling to obtain a more robust model. Nine A-domains were modeled in SASREF (26), using the C2 domain from the α2(VI) chain (12) and fitted the experimental data with a mean discrepancy factor x of 2.1 (n = 14) (Fig. 3, *A* and *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 (Fig. 3, *C* and *D*). Using reference-free methods, a three-dimensional reconstruction was calculated from 1776 particles using angular reconstitution. The three-dimensional model generated is similar in overall shape to the models produced from the SAXS analysis (Fig. 3*E*), in particular to the rigid body model,



B

[Zn^{2+}] 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

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 (*RI*) 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.

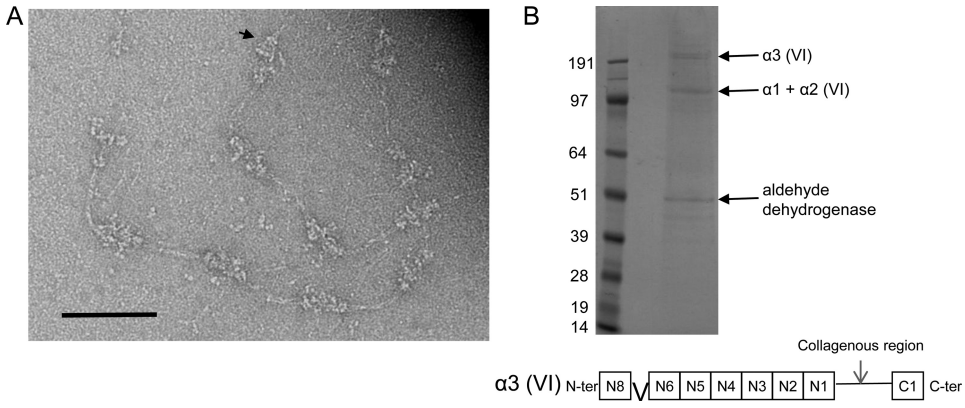


FIGURE 5. **Extraction and negative staining EM of bovine corneal collagen VI microfibrils.** *A*, electron micrograph image of a field of collagen VI microfibrils stained with uranyl acetate, taken on the FEI Polara at 200 kV, magnification ~50,000×. The image was recorded on a Gatan Ultrascan 4000 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 Fig. 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.

providing further support for a compact C-shaped conformation.

Metal Ion Binding of 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 α3(VI) chain of collagen VI includes one conserved MIDAS motif (supplemental Fig. S2) we investigated whether the α3(VI) N9-N1 region 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 μM $MgCl_2$, $MnCl_2$, or $ZnCl_2$. With the addition of 500 μM $MgCl_2$ or $MnCl_2$ there was a small decrease in the thickness of the protein layer (~0.2 nm). With the addition of 500 μM $ZnCl_2$, there was

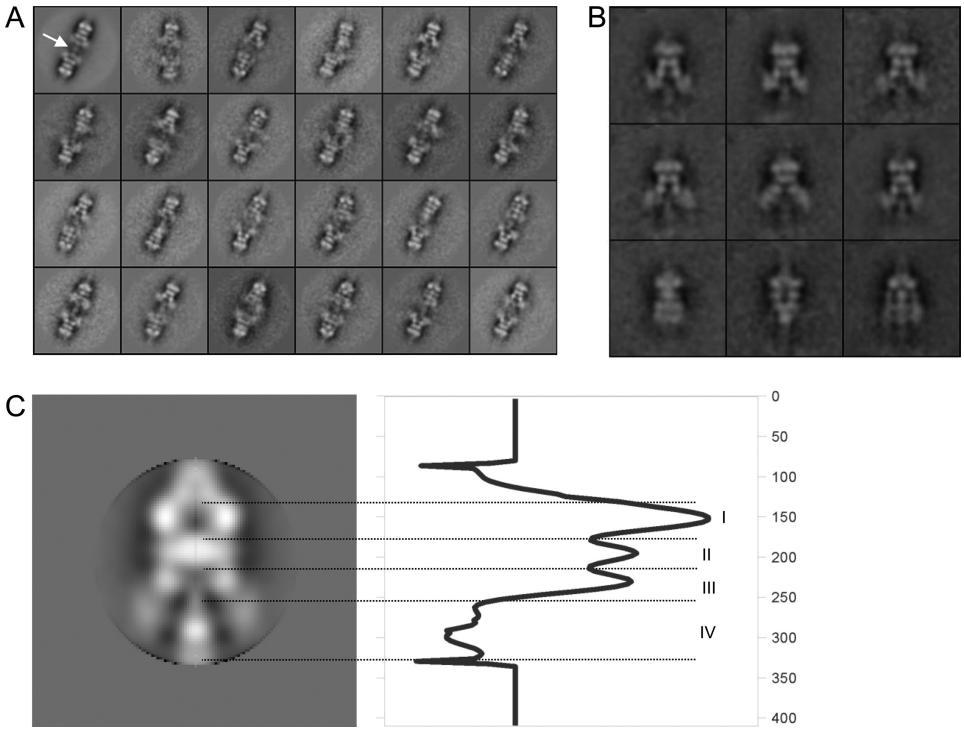


FIGURE 6. **Three-dimensional reconstruction of the half-bead region from collagen VI microfibrils.** *A*, the average of all collagen VI "double-beads," followed by 23 representative class averages. 23 of 50 class averages are shown, representing 1055 of 2438 images. The box size is 64 × 64 nm. The double-bead has approximate dimensions of ~15 × 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 nine representative class averages of images from the single particle refinement, the box size is 32 × 32 nm. Images were recorded on FEI Tecnai T12 twin at 120 kV. Some of the views can be recognized in Fig. 5. A full set of class averages and corresponding projections of the structure is shown in supplemental Fig. S3. *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 three-dimensional density (*left*) when projected to a one-dimensional trace (*right*) displays a distinctly banded structure. They have been assigned and labeled from the *top* as bands *I* (head), *II* (intermediate), and *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).

a larger decrease in thickness of ~0.8 nm (Fig. 4*A*). 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 μM) (Fig. 4*B*).

Conformational Changes in N9-N1 Region in 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 1 mM Mg^{2+} or 1 mM Mn^{2+}, the protein sediments as a single species with a sedimentation coefficient of 7.03 S (Fig. 4*C*). 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 (Fig. 4*C*), potentially representing an oligomer.

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 ±0.5%. However, in the presence of Zn^{2+}, two species are seen, one eluting at ~13 ml and another at ~15 ml (Fig. 4*D*). The species eluting at 15 ml was confirmed as a monomer (molecular mass, 197 kDa ±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 mass of 435 kDa ±0.8% (Fig. 4*D*), 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.

Three-dimensional Structure of Collagen VI Tissue Microfibrils—To determine the organization 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 V_o fraction. The major component was bovine collagen VI (α3, α2, and α1 chains detected); however, there were some other intracellular and extracellular components (supplemental Table S1). The only microfibril associated proteins detected were βig-h3 and decorin. The α6 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 (Fig. 5*A*). 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 × 64 nm boxes, and aligned by single particle methods. In the average of

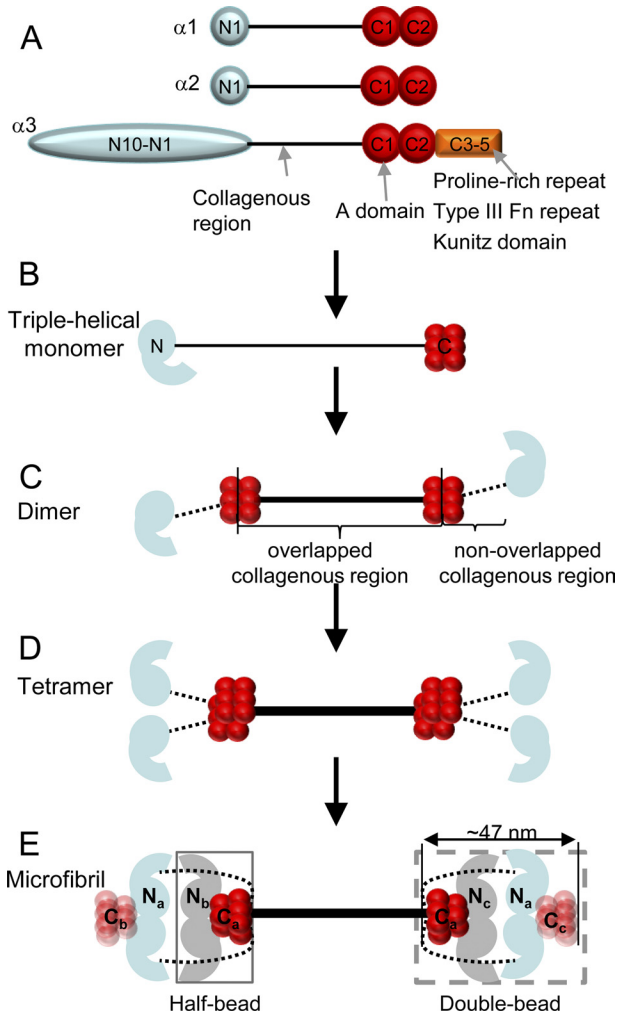


FIGURE 7. **Schematic diagrams showing the assembly of collagen VI.** *A*, domain organization of the three α-chains. *Blue* and *red circles* represent N- and C-terminal VWA domains, respectively. *B*, the α1, α2, and α3 chains assemble to form a triple-helical monomer. *C*, dimers are formed by the 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 (*gray*), C_b, and 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 Ref. 6).

the aligned bead region and in class averages (Fig. 6*A*), 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 (supplemental Movie S1). Because the collagen VI tetramer is 2-fold symmetric, a dimer of dimers, each half-bead should be an equivalent structure (Fig. 7). Therefore, for further analysis, the images of the half-beads were extracted and aligned (Fig. 6*B* shows selected class averages from the single particle refinement).

A three-dimensional map of the half-bead was computed and shows that this structure is composed of four distinct layers (Fig. 6*C*). 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 Fig. 6*C*. The intermediate region forms layer II, whereas the tail region is comprised of layers III and IV.

The head and intermediate region are compact, whereas the tail region is more variable. The head layer is slightly oval, but approximately round with diameter of 9.6 nm, whereas the intermediate region is more rectangular in shape and measures ~6.5 × 8.6 nm (Fig. 8*A*). The diameter of the tail region is in the range of 6.0 to 15.3 nm. The dome-shaped head region is hollow (Fig. 8*B*) and is fairly homogeneous, as can be seen in the averages in Fig. 6*A*. 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 three-dimensional nanostructure of nine contiguous A-domains from collagen VI. All of these techniques describe a compact arrangement of domains and converge on a single shape. This region has a compact "C"-shape with dimensions 18 × 6 × 5.5 nm (Fig. 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 two-dimensional image data, a zig-zag conformation, and a folded conformation (16). Our three-dimensional 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 that 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 (supplemental Fig. S4), which however, did not result in a conformational change as analyzed 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 dimerize in the presence of Zn^{2+} ions. The intermolecular interactions between α3 N-terminal domains could reflect associations between dimers in microfibril formation (see Fig. 7

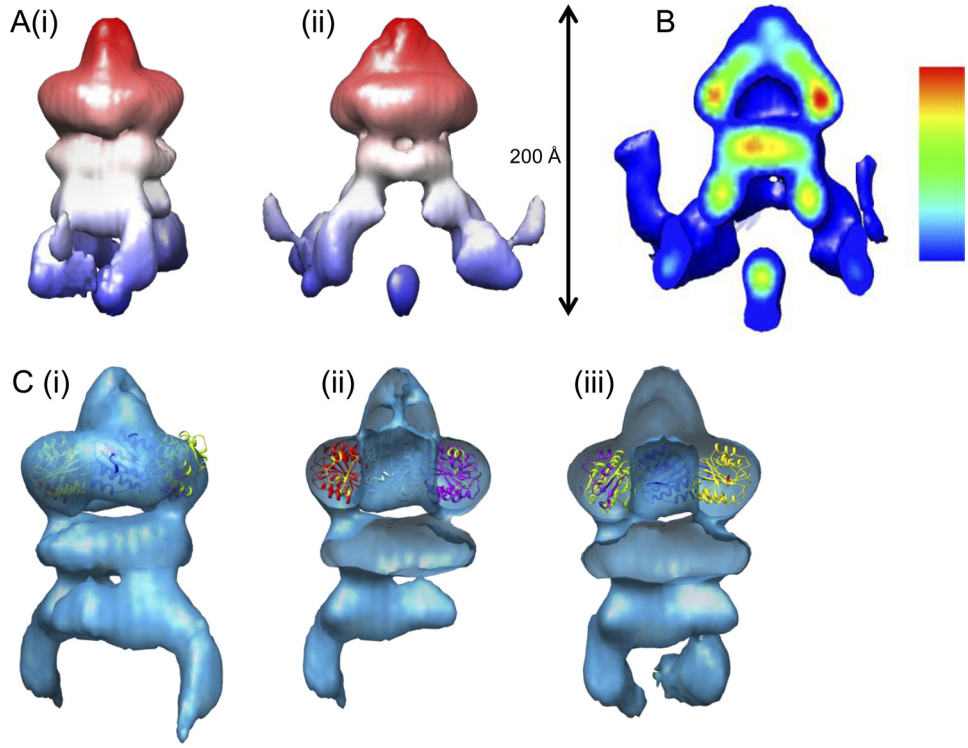


FIGURE 8. **Modeling VWA domains into the half-bead structure.** *A*, external views of the three-dimensional map are shown in two different orientations (*panels i* and *ii*). The surface is colored along the microfibril axis, *red* (head) to *blue* (tail), corresponding to assignment of the C and N termini of collagen VI (see text for explanation). *B*, cut open view of the three-dimensional map colored by relative density of the three-dimensional map. The *color key* indicates increasing density from *blue* to *red*. Maps are rendered with UCSF Chimera. *C*, three-dimensional map of the half-bead with six VWA domains fitted to the head part of the structure. *Panel i*, six VWA domains, representing the C1 domains from two triple-helical monomers, were docked into the three-dimensional map of the half-bead. *Panels ii* and *iii*, cut open views of the three-dimensional map showing three 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.

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 four 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 three-dimensional 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 (Fig. 6*C*). Measurements of the half-bead are compatible with the overlapping interaction of tetramers to form the microfibril (see Fig. 7 for explanation). This is the model most recently described in Ref. 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 α3 chain contains up to 10 VWA domains, making the N-terminal region more variable than the C terminus. The C terminus of the α3 chain has three 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 (supplemental Fig. S5). In the α3 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 three peptides were detected. No peptides were detected after residue 2496 in the C1 domain of the α3 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 α1 and α2 chains suggesting that processing only affects the α3 chain. The molecular weight estimated from SDS-PAGE is consistent with an α3 chain composed of N8-N6-C1 (see Fig. 5*B*).

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 (Fig. 8*C*). 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 α1 and α2 chains in this layer, assuming cleavage of the α3 chain C2-C5 domains. This region is possibly where N- and C-terminal domains interact, and the

layers below this (in Fig. 8) are likely to be composed of N-terminal domains.

The collagenous region runs into the head region of the half-bead. The hollow head indicates that the collagenous region must take an external path around the bead and that the globular domains are not organized around a central collagenous core. A strong feature present between the two half-beads, highlighted in Fig. 6*A*, is likely to be the collagenous region. A protrusion from the outer parts of the tails (Fig. 8, *A* and *B*) could be due to the collagenous 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 Fig. 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 α3 chain are required and a minimal N1-N5 appears necessary to support microfibril formation (15). In addition, the recently discovered α4, α5, and α6 chains all have an extended N terminus and probably assemble with the shorter α1 and α2 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.

Acknowledgments—We thank Elizabeth Lord for excellent technical assistance. We acknowledge Dr. David Knight, Marge Howard, and Emma Keevill in the Biomolecular analysis facility, Adam Huffman in the Bioinformatics facility and the staff in the EM facility, and in particular, Drs. David Holmes and Tobias Starborg (University of Manchester). We thank Professor Mon Li Chu (Thomas Jefferson University) for collagen VI DNA. We also thank Daresbury SRS for beamtime and Dr. J. G. Grossmann for assistance during data collection. The FEI Polara was supported by Wellcome Trust Grant WT081406MA.

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., Mörgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F., and Lamandé, 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. N.Y. 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. N.Y. 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. Biol. Chem.* **276,** 7422–7430
13. Lamandé, S. R., Mörgelin, 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 Lamandé, S. R. (2010) *J. Biol. Chem.* **285,** 33567–33576
15. Fitzgerald, J., Mörgelin, M., Selan, C., Wiberg, C., Keene, D. R., Lamandé, 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., Lamandé, S. R., Oldberg, A., Timpl, R., Mörgelin, M., and Heinegård, D. (2001) *J. Biol. Chem.* **276,** 18947–18952
20. Wiberg, C., Klatt, A. R., Wagener, R., Paulsson, M., Bateman, J. F., Heinegård, D., and Mörgelin, M. (2003) *J. Biol. Chem.* **278,** 37698–37704
21. Pöschl, 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. U.S.A.* **106,** 8561–8566
23. Semenyuk, A. V., and Svergun, D. I. (1991) *J. Appl. Crystallogr.* **24,** 537–540
24. Svergun, D. I. (1999) *Biophys. J.* **76,** 2879–2886
25. Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **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., Ladjadj, 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., Ylöstalo, 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., Trüeb, 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., Mörgelin, M., and Heinegård, D. (1998) *J. Biol. Chem.* **273,** 20397–20403
39. Aigner, T., Hambach, L., Söder, S., Schlötzer-Schrehardt, U., and Pöschl, E. (2002) *Biochem. Biophys. Res. Commun.* **290,** 743–748
40. Mayer, U., Pöschl, E., Nischt, R., Specks, U., Pan, T. C., Chu, M. L., and Timpl, R. (1994) *Eur. J. Biochem.* **225,** 573–580
41. Trüeb, 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