Structure of a collagen VI α3 chain VWA domain array: adaptability and functional implications of myopathy causing mutations

Received for publication, June 22, 2020, and in revised form, July 16, 2020 Published, Papers in Press, July 21, 2020, DOI 10.1074/jbc.RA120.014865

Herimela Solomon-Degefa[1], Jan M. Gebauer[2,‡], Cy M. Jeffries[3,‡], Carolin D. Freiburg[1], Patrick Meckelburg[2], Louise E. Bird[4,5], Ulrich Baumann[2], Dmitri I. Svergun[3], Raymond J. Owens[4,5], Jörn M. Werner[6], Elmar Behrmann[2,7], Mats Paulsson[1,8,9], and Raimund Wagener[1,8,9,*]

From the [1]Center for Biochemistry, Medical Faculty, University of Cologne, Cologne, Germany, [2]Institute of Biochemistry, University of Cologne, Cologne, Germany, [3]European Molecular Biology Laboratory, Hamburg Unit, Hamburg, Germany, [4]The Research Complex at Harwell, Rutherford Appleton Laboratory Harwell, Oxford, United Kingdom, [5]Structural Biology Division, Wellcome Human Genetics Centre, University of Oxford, Oxford, United Kingdom, [6]School of Biological Sciences, University of Southampton, Southampton, United Kingdom, [7]Max Planck Research Group Structural Dynamics of Proteins, Center of Advanced European Studies and Research (caesar), Bonn, Germany, [8]Center for Molecular Medicine (CMMC), Cologne, Germany, and [9]Cologne Center for Musculoskeletal Biomechanics (CCMB), Cologne, Germany

Edited by Wolfgang Peti

Collagen VI is a ubiquitous heterotrimeric protein of the extracellular matrix (ECM) that plays an essential role in the proper maintenance of skeletal muscle. Mutations in collagen VI lead to a spectrum of congenital myopathies, from the mild Bethlem myopathy to the severe Ullrich congenital muscular dystrophy. Collagen VI contains only a short triple helix and consists primarily of von Willebrand factor type A (VWA) domains, protein–protein interaction modules found in a range of ECM proteins. Disease-causing mutations occur commonly in the VWA domains, and the second VWA domain of the α3 chain, the N2 domain, harbors several such mutations. Here, we investigate structure-function relationships of the N2 mutations to shed light on their possible myopathy mechanisms. We determined the X-ray crystal structure of N2, combined with monitoring secretion efficiency in cell culture of selected N2 single-domain mutants, finding that mutations located within the central core of the domain severely affect secretion efficiency. In longer α3 chain constructs, spanning N6-N3, small-angle X-ray scattering demonstrates that the tandem VWA array has a modular architecture and samples multiple conformations in solution. Single-particle EM confirmed the presence of multiple conformations. Structural adaptability appears intrinsic to the VWA domain region of collagen VI α3 and has implications for binding interactions and modulating stiffness within the ECM.

Collagen VI is a ubiquitously expressed beaded microfibril forming collagen and serves to anchor large structures such as blood vessels and nerves to the surrounding ECM and to basement membranes (1–4). It is often found associated with other ECM components, such as decorin and biglycan. These link the collagen VI microfibrils to matrix adaptor proteins, such as matrilins (5). Collagen VI has been implicated in a wide range of physiological processes, from tumor growth and metastasis to macrophage recruitment (6). Its role in the skeletal muscle is the best characterized, as defects in collagen VI manifest as congenital myopathies (7). The hallmark conditions are Bethlem myopathy (BM) (8, 9) at the milder end of the spectrum and Ullrich congenital muscular dystrophy (UCMD) at the more severe end (10, 11).

Collagen VI was long thought to be composed of only three distinct polypeptide chains, the shorter α1 and α2 chains and the longer α3 chain. However, three alternative long chains, α4, α5, and α6, were subsequently discovered and sometimes replace the α3 chain (12, 13). In humans, the *COL6A4* gene coding for the α4 chain is inactivated by a large pericentral inversion (12). Two features make collagen VI unique within the collagen protein family. First, it contains only a short collagenous domain and, therefore, was named short-chain collagen in early studies (14, 15). Indeed, collagen VI, in particular the long chains, mainly consists of tandem arrays of von Willebrand factor type A (VWA) domains (Fig. 1). These are well-known protein–protein interaction modules of ∼200 amino acid residues. They adopt a Rossman fold of a doubly wound, open twisted central β-sheet surrounded by six α-helices (16). For example, the six murine collagen VI chains contain 46 different VWA domains and, therefore, collagen VI is the prototypical VWA domain containing multiprotein assembly. Second, collagen VI forms very large complexes of about the size of a ribosome already before secretion. The α1, α2, and one of the long chains (α3 in most tissues) form a heterotrimeric triple helical monomer (17). For formation of larger complexes to occur, two monomers assemble to a dimer in an antiparallel manner by forming a triple-helical segmented twisted supercoil (18). Two dimers then assemble laterally to form a tetramer. Several disulfide bridges stabilize the complex (1). The final secreted collagen VI tetramer, when made up by the α1α2α3 chains, contains up to 72 VWA domains. After secretion, intercalating overlapping end-to-end association of tetramers leads to the formation of the beaded microfibril, a process that is accompanied by proteolytic processing of the globular C-terminal part of the α3 chain (19, 20).

This article contains supporting information.
‡These authors contributed equally to this work.
*For correspondence: Raimund Wagener, raimund.wagener@uni-koeln.de.

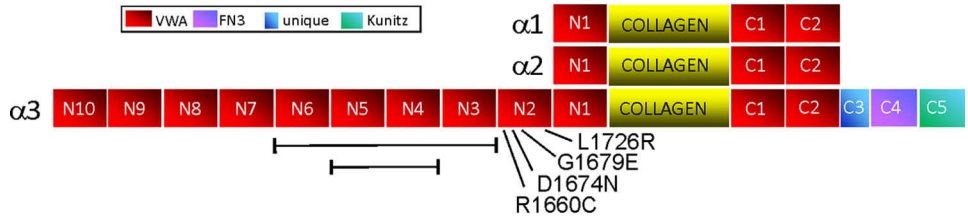

Figure 1. Schematic depiction of the domain arrangement of the collagen VI α1, α2 and α3 chains. The location of patient derived point mutations in the α3N2 VWA domain, and the tandem domains used in SAXS are indicated.

Because of its size and heterogeneity, structural studies of collagen VI at atomic resolution are challenging. A nanostructure of the bead regions from the isolated α1α2α3 chain containing collagen VI microfibrils was determined using cryo-TEM (21). The beads have a compact hollow head region with four lobes, probably containing the C-terminal VWA domains from the three α-chains. The head is connected by an intermediate region to two tail regions, most likely containing the N1 VWA domains of all three α chains. However, the tandem array of VWA domains from the N terminus of the long α3 chain is lacking from the structure (21). Nevertheless, by utilizing SAXS, a C-shaped form of the N-terminal VWA domain tandem array of the α3 chain was determined (20), a shape also found for the homologous α4, α5, and α6 chains (22). Structural information at high resolution for the VWA domain array is lacking, even though the X-ray crystal structure of the single N5 VWA domain of the α3 chain revealed a C-terminal linker extension that allows the tandem arrangement of multiple VWA domains (23).

As much of collagen VI is made up by the N-terminal tandem VWA domain array of the α3 chain, insight into its structure is required to understand collagen VI self-assembly into microfibrils and integration into the extracellular matrix. Further, such information will help in studies aimed at revealing the pathomechanisms underpinning collagen VI myopathies. Although mutations that affect the triple helix are the most common (24), several mutations in the N- and C-terminal VWA domains also have been shown to cause disease (7), and the N2 VWA domain of the α3 chain especially harbors several pathogenic point mutations (25–28). Here, we determine the structure of the collagen α3 N2 domain by X-ray crystallography and study the effects of introducing BM and UCMD patient mutations (Fig. 1) on the secretion efficiency of the single domains in cell culture. We investigate longer double and quadruple arrays of α3 chain-derived VWA domains (N5-N4 and N6-N3) by SAXS to assess the structural heterogeneity and the modular adaptability of the protein in solution. We also study the flexibility of N6-N3 by single-particle EM.

Results

The X-ray crystal structure of the collagen VI α3 N2 VWA domain reveals a pronounced structural conservation among N-terminal α3 chain VWA domains

The N2 domain located within the N-terminal region of the α3 chain carries several myopathy-causing mutations. To date, the location of the affected amino acid residues could only be estimated by homology modeling (29), and conclusions on the consequences of mutations for the pathomechanism lack an experimentally determined structural basis. Therefore, the X-ray crystal structure of the human collagen VI α3 N2 domain was determined at a resolution of ~2.3 Å, refined to R_{Work}/R_{Free} of 25.8/20.0% (Table 1). The data collected were anisotropic, with resolution estimates between 2.20 and 2.58 along different axes. This explains the relatively high R_{merge} values in the outer shell (2.27–2.20 Å). The asymmetric unit contains two independent copies of the VWA domain, which are virtually identical (all-atom root mean square deviation [RMSD] of 0.23 Å) even in the loop sections, which indicates a certain structural rigidity.

Table 1
Dataset for the N2 VWA crystal and refinement

Parameter	Value(s) for collagen VI
Data collection	
Space group	P1
Cell dimensions	
a, *b*, *c* (Å)	36.64, 40.54, 57.11
α, β, γ (°)	83.34, 76.80, 79.56
Resolution (Å)	55.43–2.2 (2.27–2.2)*
R_{sym} or R_{merge}	0.22 (1.24)*
$I/\sigma I$	4.6 (2.4)*
Completeness (%)	96.5 (96.5)*
Redundancy	3.4 (3.2)*
Refinement	
Resolution (Å)	55.43–2.2
No. reflections	15,370 (1562)*
No. of reflections used for R_{free}	761 (76)*
R_{work}/R_{free} (%)	20.0/25.7
No. of atoms	3019
Protein	2915
Water	104
B-factors	
Protein	26.86
Water	29.74
Ramachandran	
Favored (%)	96.0
Allowed (%)	3.5
Outliers (%)	0.5
RMSD	
Bond lengths (Å)	0.006
Bond angles (°)	0.65

* Values in parentheses are for highest-resolution shell.

The structure of the human collagen VI α3 N2 domain (Fig. 2*A*) is typical of the VWA domain family, adopting an alternating α/β Rossman fold made up of a central twisted core of five parallel (β1,2,4,5,6) and one anti-parallel (β3) β-strand surrounded by six α-helices. The sequence of the N2 domain does not indicate a metal ion-dependent adhesion site (MIDAS) motif, often present in VWA domains (30).

Overall, the structure of the N2 domain is very similar to that of the α3 N5 domain (PDB entry 4IGI [23]) (Fig. 2*B*) with an RMSD Cα of 0.90 Å, even though the domains share only 31% sequence identity. Remnants of the C-terminal linker (as seen in Protein Data Bank accession no. 4IGI) that would otherwise

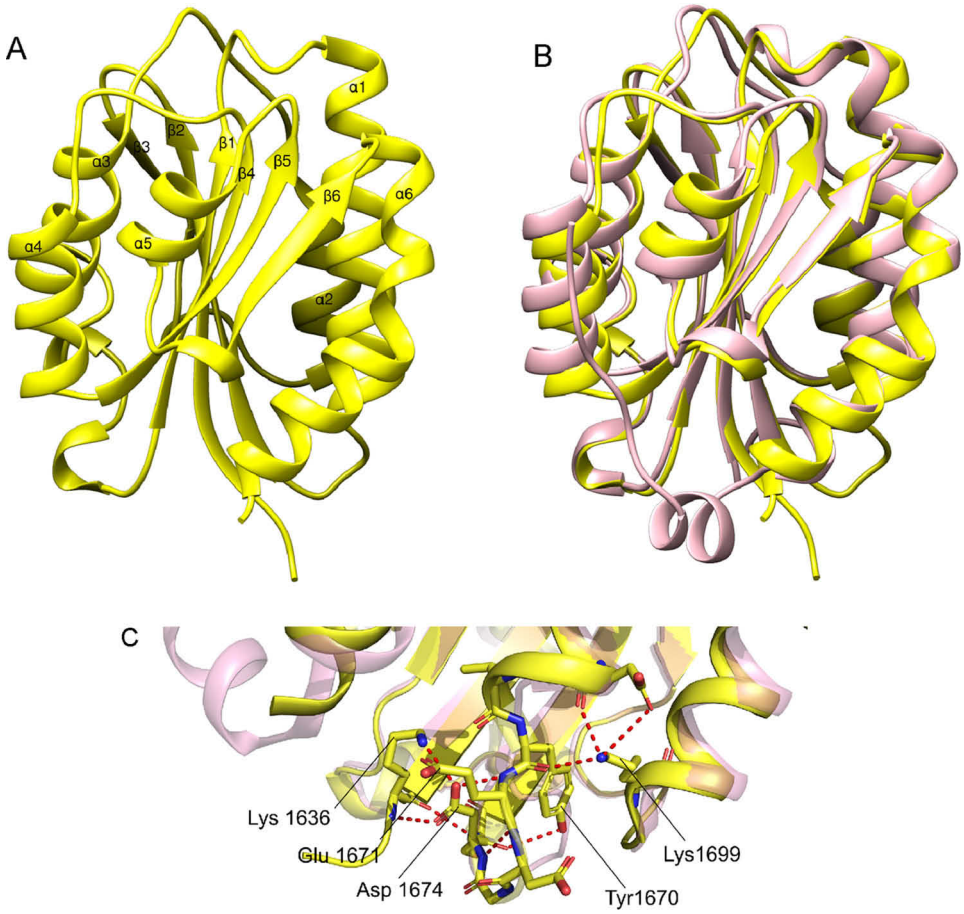

Figure 2. Crystal structure of human collagen VI α3 chain N2 domain. *A*, crystal structure solved at ~2.3 Å. α helices and β strands are numbered 1–6, based on the naming convention of the von Willebrand factor A2 domain. *B*, structural alignment of N2 domain (*yellow*) with murine collagen VI α3 chain N5 domain (PDB entry 4IGI [23]) (*pale pink*) showed a total RMSD of 0.90 Å. The only differences are in the N-terminal extension of the N5 domain and in the loop region connecting β2/α2, shown in detail in *panel C* with a hydrogen bond between the N-terminal Lys_{1636} and Lys_{1699}.

connect to the N1 domain were not resolved, suggesting a level of structural heterogeneity in this region of the protein. Besides the unresolved C terminus, the major difference between the two structures is the short loop/turn in the N2 domain between β2 and α2 with the sequence $_{1670}$YEDGD$_{1674}$. This is partly explained by the difference in primary structure in the N5 domain ($_{1055}$LDVGPD$_{1060}$), which is slightly longer and seems to allow different hydrogen bridging with Lys_{1636} at the N terminus and Lys_{1699} in the α2 helix of the N2 domain (Fig. 2*C*).

The N2 VWA domain harboring BM and UCMD mutations exhibit different self-association propensities and affect cellular secretion levels

The X-ray crystal structure of the N2 domain allows exact mapping of pathogenic mutations (Fig. 3*A*) and thereby gives clues for pathomechanisms. The mutations R1660C, D1674N, G1679E, and L1726R were selected from literature or from the Leiden Open Variation Database (LOVD; https://www.lovd.nl/), using the criteria that they were (i) likely pathogenic and/or (ii) had supporting data from patient material indicating pathogenicity (25–28). The mutations cover a spectrum of prototypic locations within a VWA domain. G1679E is located in the central hydrophobic β sheet, L1726R in an α helix facing the hydrophobic core, and R1660C and D1674N are surface exposed in an α helix and a loop, respectively. As the recombinant expression of heterotrimeric collagen VI has not yet been achieved, only the mutated single VWA domains were expressed in HEK-293 EBNA cells to study their impact on expression and secretion (Fig. 3, *B* and *C*).

The WT N2 domain is expressed and secreted and is well resolved by SDS-PAGE under nonreducing (Fig. 3*B*) and reducing (Fig. 3*C*) conditions. The N2 domain carrying the R1660C mutation (*lane 2*) is expressed and secreted into the cell culture supernatant. However, in the cell lysate, an additional faster-migrating band is seen under nonreducing conditions. The N2 domain carrying the D1674N mutation (*lane 3*) is the only mutated domain that is expressed and secreted in a manner comparable to that of the WT N2 domain. Interestingly, under nonreducing conditions, the WT N2 domain and the R1660C and D1674N mutant domains in the cell culture supernatant show evidence of disulfide bond-mediated dimerization, with a slower-migrating band within the cell culture supernatant at 55 kDa (Fig. 3*B*) that disappears after reduction (Fig. 3*C*). R1660C shows qualitatively more of this dimer than D1674N and the native N2, which shows only trace levels. The surface-exposed location of the R1660C mutation likely increases the propensity of the domain to form disulfide-bonded dimers under oxidizing

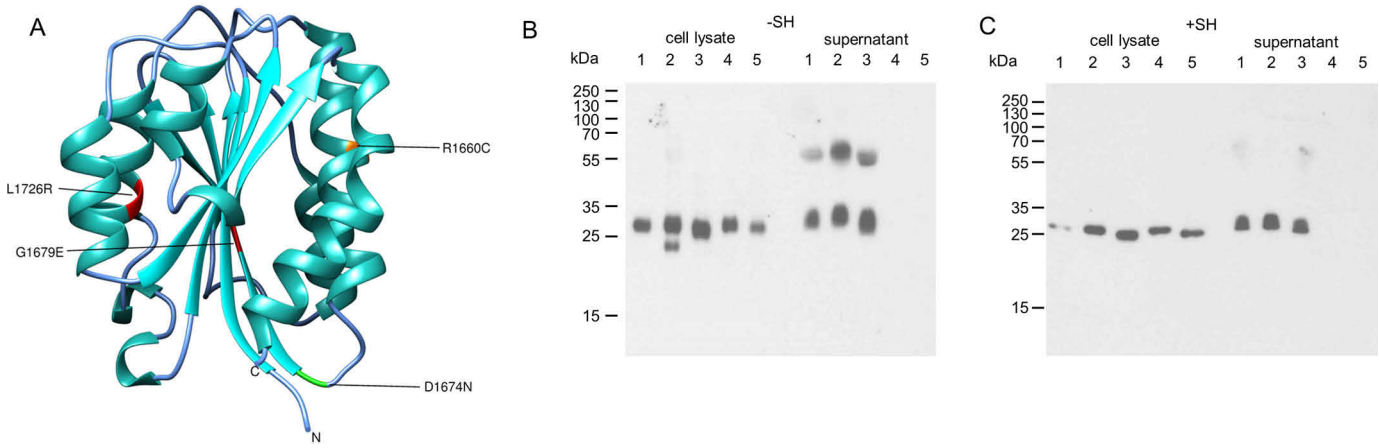

Figure 3. Location and consequences of point mutations in the collagen VI α3 chain N2 VWA domain. *A*, the structure of the N2 domain, with the location of myopathy patient mutations. Mutations are colored according to their location, either in the core (*red*), on the surface on an α helix (*amber*), or on a loop (*green*). *B* and *C*, Western blots were performed under nonreducing (*B*) and reducing (*C*) conditions. The blots were probed with an antibody directed against the Strep tag. Samples were HEK-293 EBNA cell lysates and supernatants of the WT N2 (*lane 1*) and mutants R1660C (*lane 2*), D1674N (*lane 3*), G1679E (*lane 4*), and L1726R (*lane 5*). The corresponding Ponceau staining is shown in Fig. S1.

conditions. However, the dimerization observed in D1674N and trace dimerization in the native N2 domain can only be reconciled by the formation of a disulfide bond between two unpaired cysteine residues (C_{1825}) located within the unstructured C-terminal tail. Strikingly, the G1679E (*lane 4*) and L1726R (lane 5) mutant domains are expressed but retained within the cell, shown by the strong band in the cell lysate and the lack of detectable material in the supernatant (Fig. 3, *B* and *C*).

The double VWA domain construct N5-N4 samples a definable set of structural states

Although X-ray structures of single VWA domains may help to understand their particular function, *e.g.* by identifying binding surfaces, information about the connection of VWA domains and their relative arrangement to each other in space is also needed. Therefore, attempts at crystallizing tandem double domains were made. The variant N5-N4 was chosen as a prototype because the N5 domain, including parts of the linker to N4, has already been crystallized (23). However, the attempts to crystallize N5-N4 were unsuccessful. Therefore, in-line size exclusion chromatography SAXS (SEC-SAXS) (31) combined with parallel multiangle laser light scattering (MALLS) were used to obtain the structural parameters of N5-N4, including the molecular weight (MW), radius of gyration (R_g), maximum particle dimension (D_{max}), and probable real-space atom-pair distance distribution, or $p(r)$ profile. The SAXS results are summarized in Fig. 4 and 5, and Fig. S2 and S3, and Table S2. The majority of the protein sample elutes from the gel filtration column as a monomer with an average MW of 43 kDa (expected MW from the N5-N4 amino acid sequence, 44.3 kDa) with an R_g of 2.90 ± 0.1 nm (Fig. 4*A*). The resulting $p(r)$ profile yields a highly skewed distribution of real-space distances, indicating a rather elongated particle. The profile extends to the maximum diameter D_{max} of ca. 10 nm and is dominated by a maximum at $r = 2.6$ nm and a well-defined secondary shoulder at 5 nm (Fig. 4*B*). The former feature points to the average cross-section of the particle. The latter feature is likely caused by contributions from scattering pair distances arising from well-defined, spatially separated domains. Low-resolution models of the protein in solution were calculated using *ab initio* modeling. The *DAMMIN* (32) *ab initio* model of the N5-N4 domain pair and corresponding fit ($\chi^2 = 1.03$; CorMap [33], $p = 0.85$) are shown in Fig. 4*C*. It appears that the N5-N4 monomer forms an extended dumbbell, with two compact globular lobes. Overall, the double domain construct has dimensions of ~4 × 4 × 9.5 nm, whereas each lobe is ~4 × 4 × 4 nm, such that each are of sufficient volume, and shape to accommodate a single compact VWA domain (Fig. 4*C*).

The X-ray crystal structure of the N5 domain has been solved (PDB entry 4IGI); however, a high-resolution atomistic model of the N4 domain is not available. Therefore, a homology model of N5-N4 was generated with *I-TASSER* (34) using the amino acid sequence of N5-N4 as the input (Fig. S2). The top-scoring homology model generated by the *I-TASSER* calculations has structural similarities to the double VWA domain of the proximal thread matrix protein 1 (PTMP1), isolated from the mussel shellfish (PDB entry 4CN8) (35) (Fig. 5). The N5-N4 *I-TASSER* homology model does not agree with the SAXS data ($\chi^2 = 18$, CorMap $p = 0$, Fig. S3). Of note, the R_g of the *I-TASSER* model is far too small (2.4 nm) compared with that obtained from the experimental SAXS data (2.9 nm), indicating that the VWA domains of the predicted model are not sufficiently separated. To obtain a model of N5-N4 that fits the SAXS data, the X-ray crystal structure of N5 was combined with the *I-TASSER*-built N4 domain, and a dummy amino acid linker was generated to connect the two domains. Using the program *BUNCH* (36), the spatial positioning of the N5 and N4 VWA domains was refined against the SAXS data with the optimization of the linker conformation (Fig. 5). *BUNCH* was run several times (at least 10) to produce a cohort of rigid-body refined atomistic structures of N5-N4 that fit the SAXS data ($\chi^2 = 1.0$, CorMap $p = 0.8$; Fig. 5*A*). The separation of the N5 and N4 domains of the refined *BUNCH* model agrees with the volume (mass) distribution of the *ab initio* dummy atom bead model. The linker connecting the N5 and N4 domains is not hyperextended but does allow for the N5 and N4 domains to be distinctly separated from each

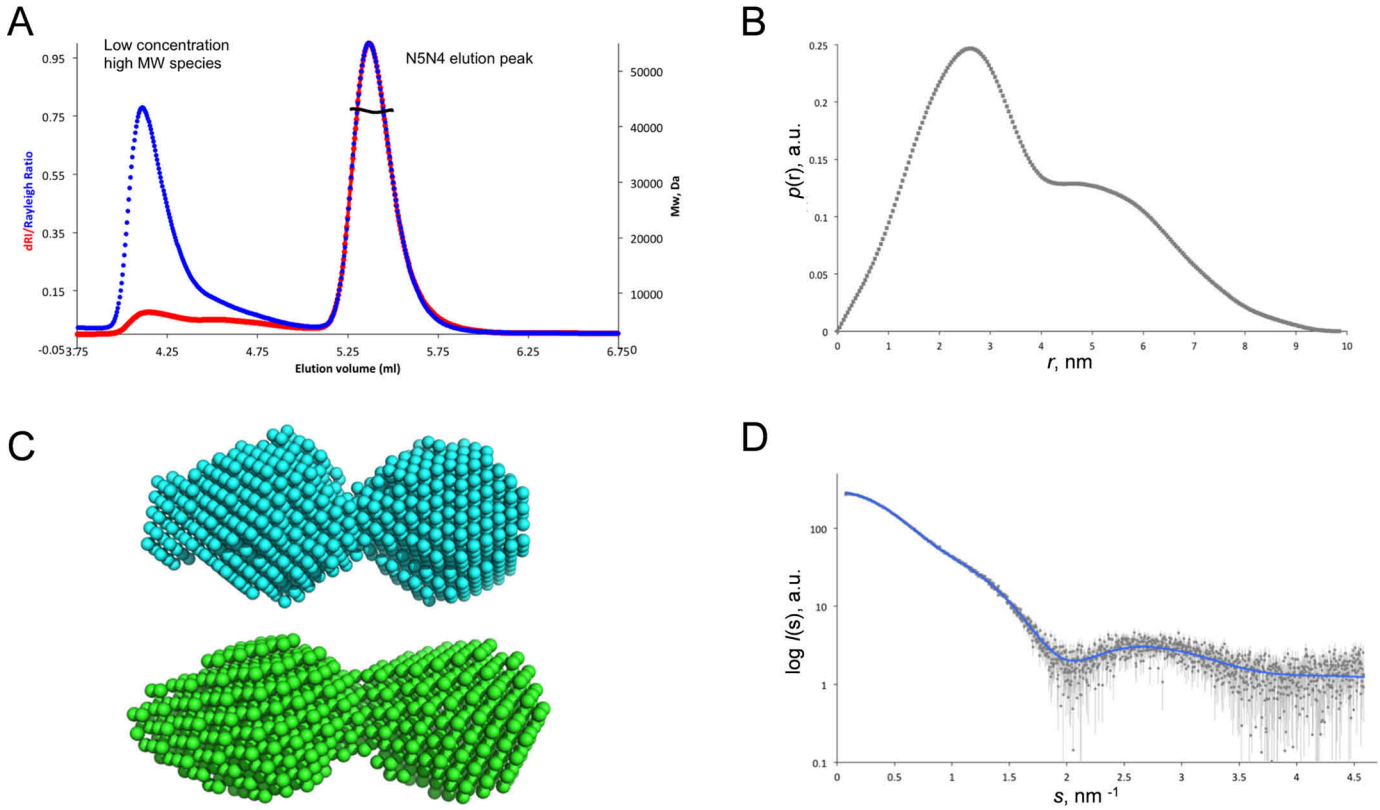

Figure 4. ***Ab initio*** **modeling of collagen VI α3 N5-N4 from SEC-SAXS combined with MALLS.** *A*, the MALLS, cm^{-1} (*blue*), and dRI, $ml \cdot g^{-1}$ (*red*), traces obtained for the gel filtration elution peak of N5-N4. The MW correlation through the N5-N4 peak is shown as a *black line* that spans 42.5–43.2 kDa (MW average, 43 kDa). *B*, the real-space scattering-pair distance distribution, or $p(r)$ profile, of N5-N4 calculated from the SAXS data [shown in *panel D*, measured as scattering intensity, $I(s)$, *versus* s, where $s = 4\pi\sin\theta/\lambda$; 2θ is the scattering angle and λ the X-ray wavelength]. *C*, *DAMMIN*-generated *ab initio* models for the N5-N4 protein showing an individual low-resolution structure (*cyan*). Ten individual *DAMMIN* models were spatially aligned and averaged, with the result shown here in *green*. *D*, the fit to the SAXS data of the individual *ab initio* model is also shown (*blue line*).

other (Fig. 5*D*). In addition, the general positioning of the N5 and N4 domains across the *BUNCH* model cohort is consistent, suggesting that the protein does not sample highly diverse structural states, although it cannot be excluded that the domains may undergo a twisting motion relative to the long axis of the protein (Fig. 5*A*, *middle*). An additional investigation to assess the structural heterogeneity of N5-N4 was performed using the ensemble optimization method, *EOM* (37, 38). The results of *EOM* analysis are shown in Fig. 5*B, C,* and *D*. Briefly, *EOM* takes the atomistic structures of the N5 and N4 domains and treats these as rigid bodies, whereas the linker between the domains (Fig. 8*D*) is treated as a flexible random chain. *EOM* then generates a cohort of 10,000 structures incorporating the random linker conformations and calculates the resulting R_g and D_{max} distributions of this initial pool of structures. Using a genetic algorithm, *EOM* finds the best subset of ensemble states to represent the experimental SAXS profile. Therefore, by comparing the R_g and D_{max} distributions of the refined pool that fits the SAXS data relative to the initial pool, it is possible to assess whether the protein samples have highly diverse states (if so, the random and refined pool distribution widths will be similar) or whether more extended or more compact ensembles best represent the conformation(s) of the protein in solution (Fig. 5*B*). For N5-N4, the final *EOM* refined ensemble ($\chi^2 = 1.04$, CorMap $p = 0.18$) has significantly narrower R_g and D_{max} distributions than the initial pool, suggesting that the protein is sampling a limited set of conformational states. In addition, the distributions shift to lower values of R_g and D_{max}, indicating that the linker connecting the domains tends toward a more compact structure (*i.e.* does not sample hyperextended conformations). The resulting bimodal nature of the R_g distribution suggests the coexistence of two states of N5-N4. The state where N5 and N4 are spatially separated, as observed from the *BUNCH* modeling, exists together with a less frequent compacted state, where the N5 and N4 domains are closer together, similar to the predicted *I-TASSER*/mussel PTMP1 model. With respect to the volume fractions of these two states, *EOM* generates a refined ensemble in which 75% of the structures are in the *BUNCH*-like extended conformation and in 25% of the cases are in a more compact PTMP1-like conformation (Fig. 5*D*). In combination, the *BUNCH* and *EOM* analyses suggest that the linker between the N5 and N4 VWA domains contributes in limiting the conformational sampling of the double domain variant, the overall structural states of which are well defined.

The quadruple VWA domain construct, N6-N3, samples a diverse conformational ensemble and is structurally heterogeneous

The four-domain N6-N3 tandem construct was also subjected to SEC-SAXS analysis with parallel MALLS measurements (Fig. 6). The MW correlation through the gel filtration

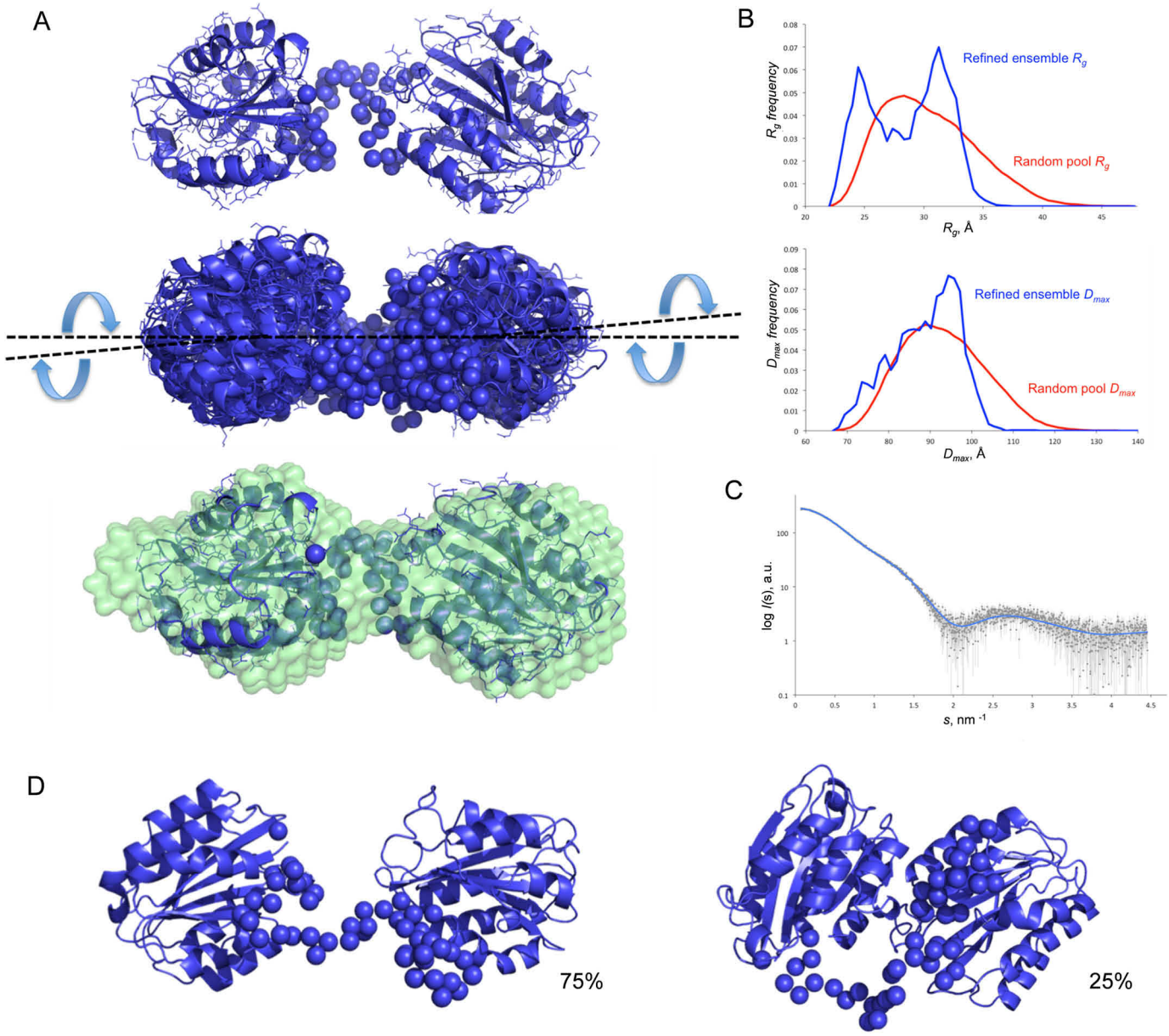

Figure 5. *BUNCH* refined rigid body model cohort and ensemble representations of N5-N4 in solution. *A*, *top*, single *BUNCH*-refined rigid body model of N5-N4; *middle*, the aligned *BUNCH* model cohort obtained from several modeling runs showing the possibilities of an axial twist of the domains relative to the long axis of the protein. *Bottom*, spatial overlay of a single *BUNCH* model and the *ab initio* model of N5-N4 obtained from *DAMMIN*. *B*, random and refined pool distributions of possible conformations of N5-N4. The R_g and D_{max} distributions obtained from *EOM* analysis comparing the initially generated random pool distributions (*red*) and the final refined ensembles of N5-N4 (*blue*) are shown. *C*, the fit of the refined pool (*blue line*) to the SAXS data. *D*, model representatives of the refined N5-N4 *EOM* pool and the estimated volume fractions of the two main structural representatives.

peak of N6-N3 is shown in Fig. 6*A*, with estimates falling within a narrow range of 85.2–90.1 kDa, with an average of 88.5 kDa. The corresponding concentration-independent MW obtained from the SAXS data were evaluated at 92 kDa. Combined, the results indicate that N6-N3 elutes from the column as a monomer (expected MW, 93 kDa).

The $p(r)$ profile of N6-N3 is shown in Fig. 6*B*. The R_g determined from the $p(r)$ at 4.25 ± 0.1 nm is consistent with the Guinier R_g estimate of 4.1 nm (Table S3 and Fig. S4). Overall, the $p(r)$ profile has two defined maxima at r = 2.8 and 5.2 nm, and the overall anisotropic distribution of distances that extend to a D_{max} of 16 nm suggests the protein has an extended/modular domain organization. Additional analysis of the SAXS data showed that the scattering profile is highly ambiguous, meaning that a number of shape topologies fit the scattering data (Table S3) (39). Therefore, the *ab initio* modeling using *DAMMIN*, and the alternative program *GASBOR* (40), were run several times, and the individual results were clustered into shape categories. The final low-resolution representations of N6-N3 are presented in Fig. 6*C* and *D*. Each reconstruction fits the SAXS data (χ^2 *DAMMIN* = 0.95–1.0; CorMap p = 0.17–0.67; χ^2 *GASBOR* = 1.0; CorMap p = 0.01–0.86), where the average spatial disposition of N6-N3 forms an extended, bent, and somewhat flattened modular domain architecture.

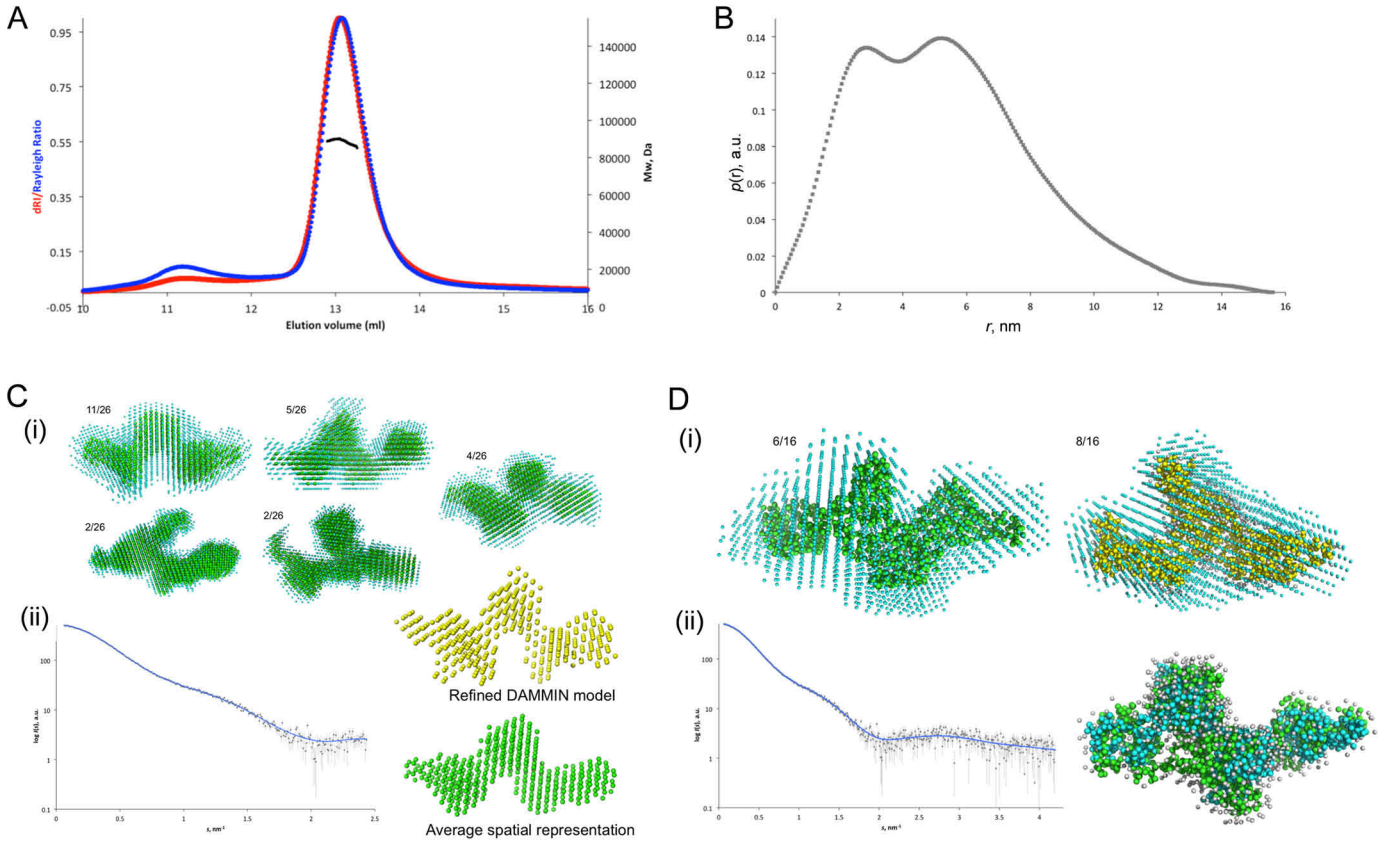

Figure 6. Generation of *ab initio* model of collagen VI α3 N6-N3 by SEC-SAXS combined with MALLS. *A*, MALLS, cm^{-1} (*blue*), and dRI, $ml \cdot g^{-1}$ (*red*), data obtained for the gel filtration elution peak of N6-N3. The MW correlation through the N6-N3 peak is shown as a *black line* that spans 85.2–90.1 kDa (MW average, 88.5 kDa). *B*, the or *p*(*r*) profile of N6-N3 calculated from the SAXS data. *C*, *ab initio* model summary and refined model of N6-N3 obtained from *DAMMIN* (*i*), with the fit to the scattering data of the models in *panel ii*. *D*, *GASBOR* clustered models highlighting the extended and somewhat modular/flattened arrangement of the N6-N3 protein (*i*) and the fit to the SAXS data of the *GASBOR ab initio* models (*ii*).

Atomistic models of N6-N3 were developed by combining the crystal structure of N5 with *I-TASSER* homology models of N6, N4, and N3 and performing both *BUNCH* and *CORAL* (41) rigid-body modeling, taking into account the missing linkers between the domains (Fig. 8*D*). The spatial positioning of the N6, N5, N4, and N3 VWA domains were refined in parallel against two SAXS datasets: the N6-N3 data and the data from the double-domain N5-N4 construct (Fig. 7*A*). Both refinement programs were run several times (at least 30) to produce a very limited set of atomistic structures that fit both the N6-N3 SAXS data (χ^2 = 1.02–1.07, CorMap p = 0.093–0.55) as well as the N5-N4 data (χ^2 = 1.03–1.12, CorMap p = 0.001–0.18). Examples of the *BUNCH* and *CORAL* models of N6-N3 are displayed in Fig. 7*B*. In general, the N6, N5, and N4 domains appear to cluster to form an trefoil-shaped head, whereas the N3 domain extends out into solution. Of note, it is difficult to ascribe one specific conformation of the protein, with the domains, and especially N3, sampling different conformational states, although these states are, on average, consistent with the mass distribution seen in the *ab initio* models. As the above analysis suggests that N6-N3 is structurally heterogeneous, *EOM* was performed against the single N6-N4 data set, and the results are summarized in Fig. 8. The refined ensemble that fits the SAXS data (χ^2 = 0.98, CorMap p = 0.9) can be represented by models where, and in general, the trefoil arrangement of N6-N4 is relatively preserved across the ensemble, although it is noted that the N-terminal N6 domain may adopt different, but spatially close, positions from those of the N5-N4 core (as is also indicated in the *BUNCH/CORAL* modeling). The apparent clustering of the N6-N4 modules is in contrast to the N3 domain that appears to sample markedly different spatial positions relative to the trefoil head. For example, 44% of the volume fraction of states within the ensemble have a spatially separated N3 domain at the end of an extended linker, whereas 55% show the N3 domain proximal to the adjacent N4 domain (Fig. 8*A*).

To corroborate the structural heterogeneity observed in the SAXS experiments, the quadruple VWA domain construct N6-N3 was directly visualized at domain resolution using negative-stain EM. Indeed, the raw micrographs highlight the structural heterogeneity of the four-domain N6-N3 tandem construct (Fig. 9*A*). To investigate the underlying conformational ensembles, we analyzed the data using both deterministic and stochastic image classification approaches. In close agreement with the SAXS data, both classification methodologies independently revealed a multitude of shape topologies, ranging from extended pearls-on-a-string conformations, with length of up to 16 nm, to compacted, square conformations with a diameter of circa 10 nm (Fig. 9*B*). Whereas an assignment of the individual domains is not possible because of their almost

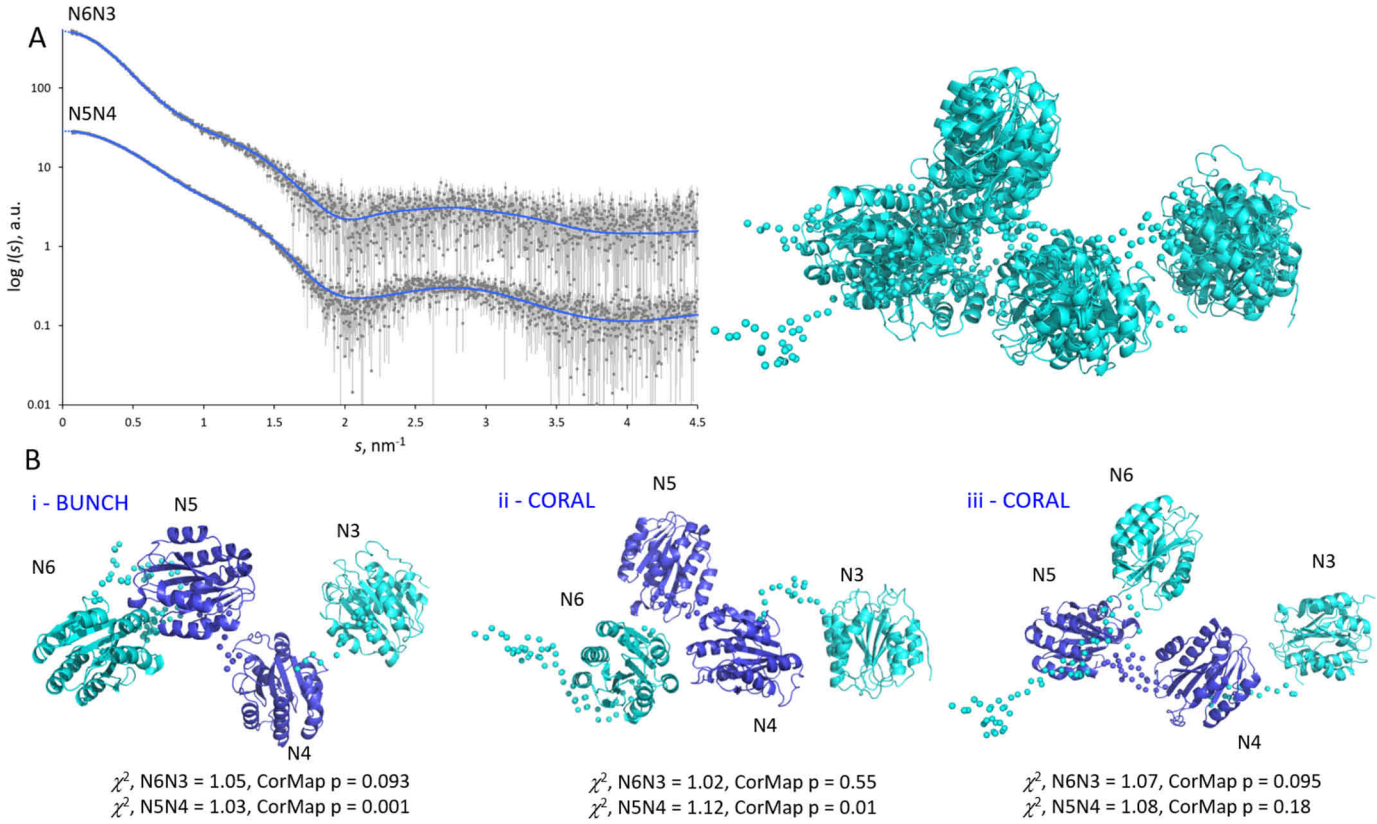

Figure 7. Rigid body models and fit to the SAXS data of VI α3 N6-N3. *A*, *left*, the fit to the SAXS data of a selected *BUNCH*-refined rigid-body model to the N6-N3 data as well as the N5-N4 core of the model to the N5-N4 data. *Right*, corresponding spatial alignment between *BUNCH* and *CORAL* generated N6-N3 structures. *B*, the three individual *BUNCH* and *CORAL* models (N5-N4, *blue*; N6 and N3, *cyan*) that fit both the N6-N3 and N5-N4 SAXS datasets. Although the N6-N4 domain region of the *CORAL* model (*iii*) appears different from *BUNCH* (*i*) and *CORAL* (*ii*), this depends on how the models are aligned (whether N3 projects away from, or into, the plane). For clarity, the N5-N4 SAXS data shown in *panel A* have been scaled by 0.1 on the *I*(*s*) axis.

identical dimensions, most topologies observed in the class averages, corresponding to circa 90% of all classified particles, agree with the SAXS data assignment of a trefoil-shaped head region with a more mobile fourth domain (Video S1). Independent of the classification methodology used, a subset of classes corresponding to circa 10% of all classified particles shows only three resolved domains (Fig. S5). Inspection of the individual particle images constituting these classes revealed that this is likely because of occlusion of the fourth domain by the other three domains.

The combined SAXS and negative-stain EM results demonstrate that structural heterogeneity and conformational sampling is intrinsic to the modular VWA domain architecture of the collagen VI α3 chain. The combined modeling results, encompassing N5-N4 and N6-N3, suggest that different levels of interdomain linker flexibility combined with possible transient/frustrated interactions between each of the individual VWA modules endows the collagen VI α3 chain with a propensity toward structural adaptability.

Discussion

Collagen VI is a very large, microfibril-forming multichain ECM protein. The three classical chains carry many VWA domains both in their N- and C-terminal regions, with the α3 chain containing the most (20, 21, 42). Here, we analyzed structural features of the large tandem array of VWA domains at the N terminus of the α3 chain. Our work focused on the myopathy mutation-prone N2 VWA domain and on the central domains N6 to N3 that are most likely relevant for the proposed flexibility of the N terminus of the α3 chain (21). These may be required for microfibril formation (43) or for the ability to bind other extracellular matrix macromolecules, *e.g.* biglycan and decorin (5) or heparin and hyaluronan (44).

The location of point mutations in the structure of the N2 VWA domain determines how the local fold is affected and indicates the severity of the resulting myopathy

Outside of the triple helical domains collagen VI is highly polymorphic (45), and for the N1-N10 domains of the α3 chain, 257 missense mutations of uncertain clinical significance are listed in the ClinVar database. Therefore, it is difficult to discriminate between polymorphisms and disease-causing missense mutations. Because exome or whole-genome sequencing has become a standard tool for the diagnosis of collagen VI related myopathies, this is an emerging problem. The lack of structural information makes it especially difficult to predict the consequences of changes in the amino acid sequence for the structure and, therefore, the function of the protein. The N2 domain of the α3 chain, on the boundary between the triple helical core region and the extended array of N-terminal VWA

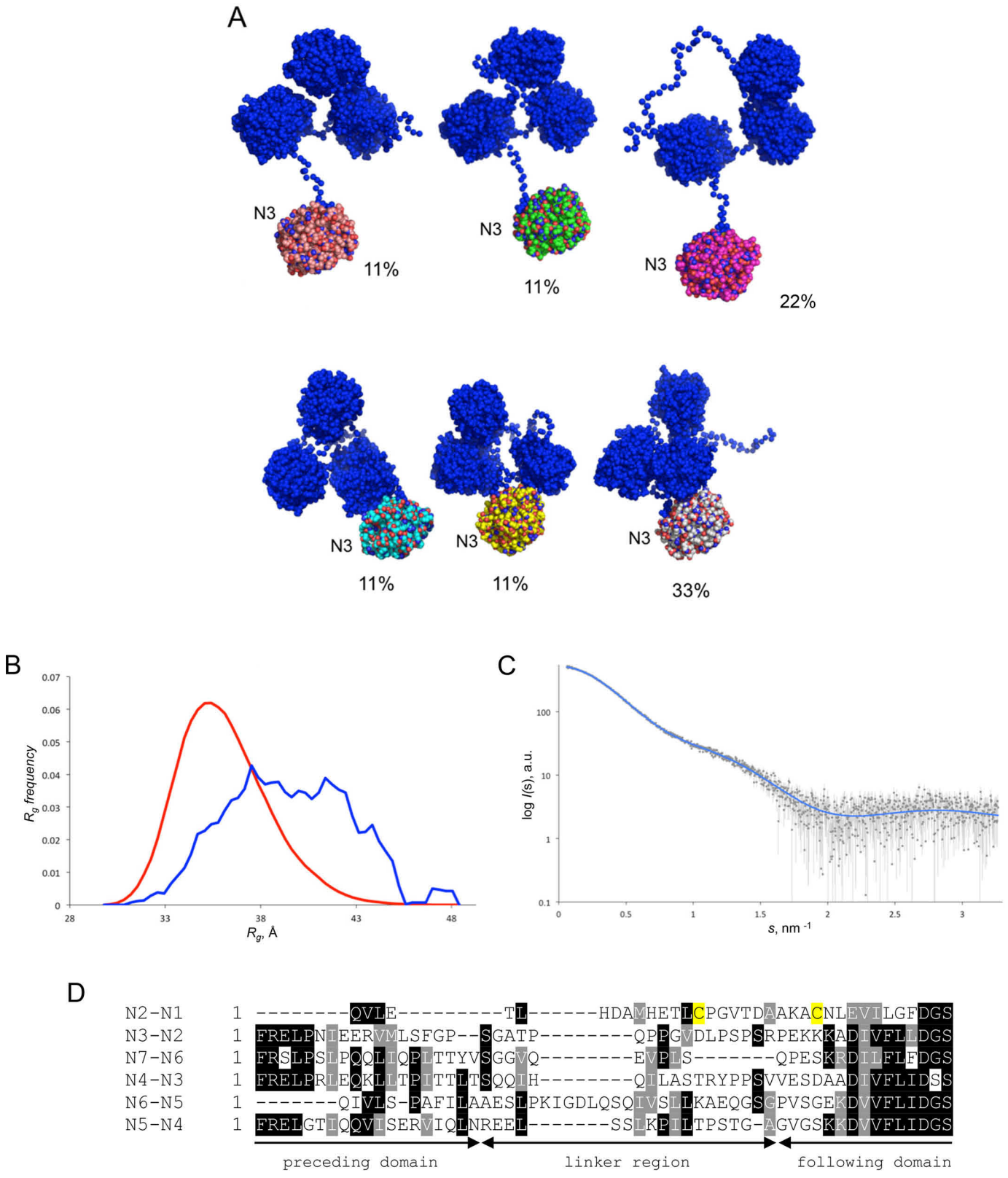

Figure 8. Summary of the EOM analysis of the N6-N3 construct and alignment of linker regions of the domains N7-N1. *A*, model representatives of the refined N6-N3 EOM pool and the estimated volume fractions of the structural representatives. The N6-N4 domains are colored *blue*, whereas the N3 domain is labeled showing that it can adopt multiple conformations that sample proximal or distant positions relative to the N6-N4 trefoil head. *B*, the R_g distribution obtained from EOM analysis comparing the initially generated random pool distribution (*red*) and the final refined ensemble R_g distribution of N6-N3 (*blue*). *C*, the fit to the SAXS data of the final ensemble (*light-blue line*). *D*, an alignment of the linker regions between the domains N7-N1. Cysteine residues are highlighted in yellow and the different regions are labeled beneath. Alignments were performed with the *ClustI Omega* alignment tool and the *Boxshade* tool.

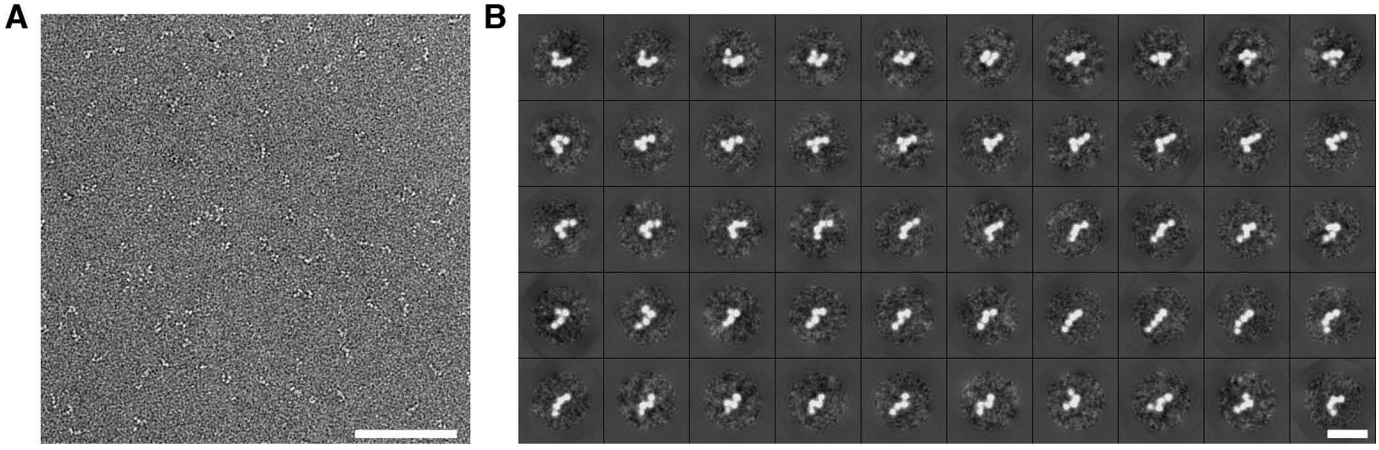

Figure 9. Transmission EM of collagen VI α3 N6-N3. *A*, excerpt of a representative TEM micrograph of negative-stained purified collagen VI α3 N6-N3. The *scale bar* corresponds to 100 nm. *B*, selected representative class averages showing the four subunits expected for the N6-N3 particles. The *scale bar* corresponds to 20 nm.

domains, harbors several missense mutations that have been proposed to be pathogenic or likely pathogenic (25–28). Notably, the α3 N2 domain has previously undergone crystallography attempts (29), which were unsuccessful. This is reminiscent of attempts to recombinantly express and maintain matrilin VWA domains, which are similar to those in collagen VI, in a soluble state for downstream characterization (46). Our initial attempts with the N2 domain also did not yield crystals. However, when we alkylated the free thiol group in the C-terminal linker region, crystals were obtained, indicating that disulfide bond formation interferes with crystallization. The alkylation will most likely not alter the structure of the N2 domain, as it does not contain a disulfide bond that connects N- and C-terminal ends, like the earlier crystallized N5 domain (23). Interestingly, the comparison of the tertiary structures of the N2 and N5 VWA domains revealed that only sharing a 31% sequence identity is sufficient to maintain the conserved structure (Fig. 2*B*), perhaps implying a functional conservation of VWA domains in the N-terminal array of the α3 chain. The low identity of the amino acid sequences, and therefore, the nucleotide sequences may be necessary to avoid homologous recombination events, especially as each of the N2 to N10 VWA domains is encoded on a single ~600-bp-long exon. Whereas the C-terminal extension could not be resolved in the crystal structure of the N2 domain, incidental SAXS analysis could model a C-terminal extension comprising the residues corresponding to the linker region between the N2 and the N1 domain (data not shown, but refer to the Small Angle Scattering Biological Data Bank, SASBDB, entry SASDJJ4).

The missense mutations occurring in the N2 VWA domain include two dominant mutations linked to BM, p.G1679E (25, 29) and p.L1726R (26), one mutation linked to an intermediate phenotype, p.R1660C (28), and one dominant UCMD-associated mutation, p.D1674N (27). The exact localization of the mutations revealed by the N2 domain X-ray structure, in concert with recombinant expression of single mutated N2 domains, enabled us to further pinpoint the pathomechanisms. The BM-causing G1679E mutation was found in a Dutch pedigree and segregated to 19 affected members (25). Fibroblasts from these patients do not show a significantly altered secretion of the collagen VI chains. However, they had a 20–30% loss of WT N2 domain epitopes, indicating either unfolding of N2 or degradation of the mutated chain (29). As predicted (29), the mutation is indeed located in the β2 strand of the central β sheet and thereby has a detrimental effect on the fold (29). This explains why only trace amounts of the single G1679E N2 domain could be purified from cell culture supernatant of transfected HEK-293 EBNA cells (29) and why the G1679E mutation caused the intracellular retention of the N2 domain (Fig. 3, *B* and *C*). Nevertheless, based on published data, it is likely that a mutated full-length α3 chain containing an unfolded N2 domain can be secreted (25). However, assembly and microfibril formation have not been studied in patient fibroblasts and on sections. As haploinsufficiency of collagen VI is not associated with a clinical phenotype (47–49), perhaps binding sites located on the N2 domain that are necessary for collagen VI assembly or for binding to other ECM molecules are affected. The other dominant BM mutation, L1726R, has been studied in patient fibroblasts and shown to have no consequences on assembly and microfibril formation (26) and most likely represents the same pathomechanism. As for G1679E, the single L1726R N2 mutant domain is not secreted (Fig. 3, *B* and *C*). The location of the mutation on the amphipathic α helix 3 points to a disturbed fold of the central hydrophobic core (Fig. 3*A*); therefore, a specific loss of function of the N2 domain in collagen VI microfibrils carrying the mutation is the common disease-causing effect.

The R1660C mutation occurred in a compound heterozygous patient with an intermediate BM/UCMD phenotype (28). On the paternal allele, he carried a premature termination codon, resulting in nonsense-mediated decay of the resulting mRNA. The loss-of-function mutation on the paternal allele and the R1660C mutation on the maternal allele led to a reduced level of collagen VI in the matrix of cultured fibroblasts, as well as a highly perturbed morphology of the collagen VI matrix (28). Interestingly, the mother showed signs of an undifferentiated connective tissue dysplasia, including soft scoliosis of the thoracolumbar region, skin hyperlaxity in the elbow

joint area, and hypermobility in the interphalangeal joints of the hands and feet (28). In addition, fibroblast cultures from the mother showed more intracellular accumulation and a somewhat lower level of an extracellular deposition of collagen VI (28), indicating a mild dominant effect of the mutation. The R1660C mutation is found on the surface of the domain (Fig. 3*A*). Unlike the BM-causing mutations G1679E and L1726R, which destroy the fold of the central hydrophobic core and result in intracellular retention, the R1660C N2 domain is secreted, and, under nonreducing conditions, dimerization is observed in the cell culture supernatant (Fig. 3*B*). However, although not so pronounced, the WT N2 domain already forms dimers, most likely through the formation of a disulfide bridge between the cysteine residues in the C-terminal linker region (Fig. 3*B*). Interestingly, for the R1660C N2 domain this dimerization is not observed in the intracellular fraction. Instead, a faster-migrating band occurs that is not present under reducing conditions (Fig. 3, *B* and *C*), indicating that the fold of the faster-migrating species is maintained by an intramolecular disulfide bridge between the novel cysteine and the cysteine in the N2 domain linker, leading to a more compact form of the protein (Fig. 3*B*). Compared with the mutations that affect the general fold of the N2 domain, the introduction of a cysteine residue on an exposed surface could present a novel site for aberrant disulfide shuffling within the collagen VI protein or between collagen VI and other ECM proteins.

The D1674N mutation was originally found to occur in one out of 79 patients with UCMD and did not appear in healthy controls (7). However, the genomic background of the single UCMD patient harboring the D1674N mutation is complex, as the compound heterozygous individual also carried an R1395Q mutation in the α3 chain N4 domain and was homozygous for an R876S change in the C2 domain of the α2 chain, causing a highly deleterious UCMD phenotype (27, 50). Therefore, the contribution of the D1674N mutation is difficult to assess. The recombinant expression of the N2 domain carrying the D1674N mutation does not reveal any expression differences compared with the WT in either secretion or migration pattern on SDS-PAGE (Fig. 3, *B* and *C*). The position of the mutation on the surface of the domain, coupled with the change from a negatively charged aspartic acid to an asparagine, could affect intra- or intermolecular interactions either through the gain or loss of an interaction surface (Fig. 3*A*). The consequences of this mutation can be compared with those of the R1064Q mutation, occurring in the α3 chain N5 domain, where the fold and the migration of the recombinant protein was comparable with that of the WT (23). Therefore, the lack of a significant effect of the D1674N mutation on the expression and secretion of the single N2 VWA domain is unsurprising. Rather, the data indicate that the severe phenotype observed in the patient is because of the other attendant mutations rather than the D1674N variant. Of course, the D1674N mutation may, although not being disease causing, contribute to the phenotype as a modifier compounding the effect of other mutations in the mode of closely spaced multiple mutations, which act in a coordinated fashion (51). The difficulty that the D1674N and R1064Q mutations highlights is the complexity of distinguishing genuine disease-causing variants from SNPs, which commonly occur in the genes encoding collagen VI (27). While important to understand the molecular consequences of mutations in the VWA domains, it is still difficult to dissect the effect these have on the larger α3 chain, or, indeed, on the collagen VI molecule as a whole. Limitations of studying point mutations in recombinantly expressed domains must be considered when investigating their effect, as catastrophic outcomes for the recombinant domain may be well tolerated in the context of the longer chain (26). Nevertheless, studying the consequences of mutations for folding of single VWA domains is helpful to predict pathomechanisms when patient fibroblasts are not available.

Our characterization of single point mutations in the α3 chain N2 domain offers important insights into the consequences of perturbations to the VWA domain fold. Based on these data, the design of chemical chaperones to assist in the folding of core-based mutations (G1679E and L1726R) could follow. Moreover, as the VWA domains of the α3 chain are encoded on exons with phase 1 boundaries (52), siRNA- or oligonucleotide-based knockdown or skipping of the regions coding for the VWA domains with surface mutations could be used to ameliorate the effect of aberrant interaction processes that arise from these mutations.

The structural heterogeneity of the modular VWA α3 chain could provide a mechanism for adapting to binding interfaces and modulating stiffness in tissues

The distal VWA domains of the α3 chain, upstream of N2, are rarely implicated in the pathogenesis of collagen VI myopathies (7). Instead, their roles are in the assembly of collagen VI (43) and its interactions with other ECM macromolecules, such as von Willebrand factor and heparin (44, 53). Our results show why this region has resisted interrogation using high-resolution structural methods (20–22). When combined in tandem, the VWA modules spanning the N6-N3 array (although encompassing an apparent region of increased spatial consistency across N5-N4) are not, overall, rigidly well defined but sample a cohort of positions to form an ensemble of states in solution. The flexibility of N6-N3 was also directly observed in negative-stain EM, confirming that structural heterogeneity is inherent to this part of collagen VI.

In retrospect, the inability to crystallize the tandem VWA domains described here (despite our repeated attempts) provides a hint toward this structural heterogeneity inherent to the protein. Tandem VWA domains occur also in vitrin, cochlin, and AMACO, but high-resolution structural information is not available. Each of the three VWA domains in the von Willebrand factor have been crystallized separately but not as tandem multiple-domain constructs. Of note, flexibility between the VWA domains in the von Willebrand factor may be a prerequisite for the shear-induced unfolding of the central A2 domain, making this domain accessible for ADAMTS13 cleavage (54). Interestingly, a tandem of two VWA domains from the proximal thread matrix protein 1 of the mussel byssus has been crystallized and the structure solved (35). Here, the VWA domains are connected by a two-β-stranded linker that is further stabilized by disulfide bonds, yielding a novel structural

arrangement. However, the short interdomain linkers spanning the N9-N2 array of the collagen VI α3 chain do not contain such cysteines that are otherwise present between the N1-N2 domains or in the related α4, α5, and α6 chains (Fig. 8*D*). It may be that the presence of cysteine residues, and the formation of subsequent intramolecular disulfide bonds, goes toward restricting conformational variability (23). This variability could be a contributing factor for the differential temporospatial expression of the four long chains of collagen VI. As the N6-N3 portions of the α3 chain have the capacity to sample multiple state(s), which long chain is expressed at what time could reflect a mechanism for modulating the stiffness (22) or binding site availability of the collagen VI microfibrils in a given tissue (42). For example, the α3 chain N9-N2 region has been implicated in the binding of heparin and hyaluronan (44), von Willebrand factor (53), and decorin and biglycan (5). In particular, the N8 domain has been shown to be an interaction site for von Willebrand factor (53). The alternative splicing of this domain changes the responsiveness of the collagen VI protein toward von Willebrand factor interactions, allowing the modulation of the signal cascade. Therefore, mutations in the N-terminal α3 chain VWA region may affect binding events both directly, via removal of key binding determinants and/or domain destabilization, or indirectly, by altering the conformational sampling of the protein. The intrinsic structural heterogeneity and modular architecture of the α3 VWA array and its apparently dynamic nature could provide a mechanism for stiffness modulation in tissues and provide an adaptable avidity for intra- or intermolecular interaction surfaces necessary for interacting with diverse binding partners.

The conformations of the single (N2) and tandem (N5-N4, N6-N3) α3 chain N-terminal VWA domains presented here raise interesting avenues for further investigation of this region in both WT and myopathic conditions. Such studies could ascertain to what extent conformational sampling is requisite for the correct functioning of collagen VI. In addition, it is important to dissect how pathogenic mutations, which affect the local structure of the VWA domain, result in catastrophic effects of the collagen VI protein as a whole. Such mechanisms could encompass loss of binding sites used for self-assembly or for adaptor proteins within the collagen VI interactome or simply lead to a decreased level of available collagen VI. It would be informative to identify binding surfaces involved in intra- or intermolecular interactions so as to understand the role of collagen VI both in normal tissue function and in myopathic processes.

Experimental procedures

PCR amplification of the collagen VI α3 chain VWA domains

The N2, N5-N4, and N6-N3 constructs were obtained using PCR and human/murine cDNAs as templates. Primers were designed with overhanging tags for downstream infusion cloning into infusion-enabled vectors (Table S1). All high-throughput PCR amplification was performed at the Oxford Protein Production Facility (OPPF), Rutherford Appleton Laboratories, Oxfordshire, UK. All steps were performed in 96-well PCR plate format. For the production of the mutation-containing constructs (α3 N2 R1660C, D1674N, G1679E, L1726R), the insert was designed and ordered as a geneblock from IDT technologies and used at a concentration of 20 ng/μl template for PCR with the appropriate primers. Sequences were verified by Sanger sequencing.

Recombinant expression and purification of collagen VI α3 chain VWA domains

Isolated DNA plasmids were used to transform BL21 Rosetta/BL21 Rosetta pLacI protein expression bacteria. 10 ml of a preculture was expanded to each 1-liter LB flask, and 1 mM IPTG was added to induce expression of the recombinant protein. The cultures were then incubated at 28 °C for 18 h and harvested by centrifugation. The cell pellet was lysed by sonication and PMSF added to a final concentration of 10 mM. The crude lysate was then centrifuged to separate the soluble intracellular material from the cell debris. As the N2 domain of the α3 chain contains a single cysteine, purification of this domain also included a reduction and alkylation step. 5 mM DTT was added, and after 30 min, iodoacetamide was added to a final concentration of 15 mM. An affinity chromatography column was prepared with Streptactin-Sepharose resin. The resin was washed with TBS and the supernatant of the cell lysate applied. The column was again washed with TBS and bound protein eluted with 2.5 mM D-desthiobiotin. All constructs contained either a thrombin (N5-N4) or 3C protease (N2, N6-N3) cleavage site immediately downstream of the Strep tag. Thrombin was incubated with N5-N4 at a dilution of 1 unit/1 μg protein in 20 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, pH 7.4. The N2 construct was incubated with Precission Protease (commercial 3C protease) at a dilution of 1 unit/100 μg protein in 20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8.0. The solution was dialyzed and reapplied to the Streptactin-Sepharose column, where cleaved protein was eluted in the flow-through. Excess buffer was removed by ultrafiltration. For final purification, an ÄKTA purifier FPLC system was used along with a Superdex 75 Increase 10/300 column (N2, N5-N4) or Superdex 200 Increase 10/300 column (N6-N3). The fractions contained within the UV_{280} peak were analyzed by SDS-PAGE, pooled, and concentrated by ultrafiltration.

Transient transfection of HEK-293 EBNA cells

HEK-293 EBNA cells were cultured in Dulbecco's modified Eagle medium (DMEM) with 5%, v/v, fetal calf serum and 1%, v/v, penicillin and streptomycin. 300,000 cells/well were plated into a 6-well plate and incubated at 37 °C/5% CO_2 for 18 h, followed by washing with PBS and adding fresh media. The transfection solution was prepared by incubating 500–1000 ng of the plasmids encoding the recombinant proteins with 97 μl of DMEM, 3 μl Fu-GENE Lipo-Transfection reagent for 15 min and then pipetted onto the medium layer of the cells. 24 h posttransfection, the supernatant was collected, centrifuged, and treated with 1 × Complete protease inhibitor. To harvest the cells, RIPA buffer was added and the cell layer scraped from the plate. The cells were extracted at 4 °C overnight with end-over-

end turning. After 16 h, the cells were sonicated at 30% amplitude for 10 s and centrifuged and the supernatant collected.

SDS-PAGE and Western blotting

Samples were incubated with Laemmli sample buffer with or without 5% β-mercaptoethanol and loaded into the wells, and electrophoresis was carried out. Proteins were detected either by staining with Coomassie brilliant blue R250 or by Western blotting. The proteins were electrophoretically transferred to nitrocellulose. The nitrocellulose membrane was incubated in 5% (w/v) milk powder in TBS, and a primary antibody directed against the Strep tag was added, followed by a secondary antibody horseradish peroxidase conjugate. The membrane was then incubated with ECL solution and exposed on X-ray film.

Establishment of crystallography trials

Proteins were prepared for crystallography trials by centrifugation at 13,000 rpm at 4°C for 15 min to remove air bubbles. The Mosquito crystallization robot (TTP Labtech) was initialized and a preset vapor diffusion sitting-drop program was loaded. The protein was mixed with buffer from the reservoirs of the commercial Morpheus crystallography screen to yield final buffer-to-protein ratios of 1:1, 2:1, and 1:2. Once complete, the plate was removed from the robot, sealed with an adhesive plastic seal, and stored at room temperature. Plates were checked weekly with a microscope to detect crystal formation.

Crystallization, data collection, and structure determination

Crystallization experiments were set up employing a Mosquito robot and drop sizes of 200 nl. Crystals of dimensions smaller than 50 μm × 80 μm were grown under the Morpheus D1 condition containing 2 mM oxometalates (0.5 mM sodium chromate tetrahydrate, 0.5 mM sodium molybdate dehydrate, 0.5 mM sodium tungstate dehydrate, 0.5 mM sodium orthovanadate), 0.1 M buffer system 4, pH 6.5 (MOPSO, Bis-Tris), and 50% (v/v) precipitant mix 5 (30%, w/v, PEG 3000, 40%, v/v, 1, 2, 4-butanetriol, 2%, w/v, NDSB 256) at 20 °C by vapor diffusion, cryoprotected in perfluoropolyether, and flash frozen in liquid nitrogen. Data were collected at the beamline ID30B at the European Synchrotron Radiation Facility (ESRF) in Grenoble.

The data were processed in P1 using the DIALS/DUI package (55) and the Pointless-Aimless-cTruncate (56) pipeline from within CCP4 (57). The phases were determined by molecular replacement with a VWA domain from collagen VI solved earlier (PDB entry 4IGI) as the search model using PHASER (58). The initial model was optimized by various build-and-refine cycles using COOT (59) and phenix refine (60). Further analyses and pictures were made using the program Chimera (61). The final model and data were deposited at the PDBe under the accession code 6SNK.

SAXS measurements and data reduction

SAXS experiments were carried out at the European Molecular Biology Laboratory bioSAXS-P12 beam line (DESY, Hamburg, Germany) (62). SAXS data collection, $I(s)$ *versus* s, where $s = 4\pi\sin\theta/\lambda$; 2θ is the scattering angle and λ the X-ray wavelength (0.124 nm; 10 keV), was performed at 20 °C using coupled size-exclusion chromatography (gel filtration) with a split-flow in-parallel MALLS/refractive index (RI) detector (63).

The MALLS/RI data were measured at 25 °C using a Wyatt Mini-Dawn TREOS with an in-built quasi-elastic light scattering module coupled to an OptiLab T-Rex refractometer for protein concentration determination. The MALLS system was calibrated relative to the scattering from toluene and, in combination with concentration estimates obtained from RI, was used to evaluate the MW distribution of species eluting from the gel filtration column. The dn/dc N5-N4 was set at 0.185 ml g^{-1}, whereas N6-N3 was taken as 0.1831 ml g^{-1} (calculated using SEDFIT [64]). The molecular weight estimates from MALLS/RI and the hydrodynamic radius, R_H, derived from quasi-elastic light scattering were determined using Wyatt ASTRA7 software.

All samples underwent slow defrosting from snap-frozen −80 °C aliquots, first at −20 °C for 2 h followed by thawing on ice. The sample injection volumes and load concentrations were the following: N5-N4, 55 μl at 14 mg/ml; N6-N3, 75 μl at 7 mg/ml. A GE-Healthcare S75 Increase (N5-N4) or S200 Increase 10/300 column (N6-N3) was used to affect the separation of each monomer variant. The columns were equilibrated in 20 mM Tris, pH 7.4, 150 mM NaCl, 3%, v/v, glycerol at a flow rate of 0.6 ml·min^{-1}. Glycerol was added to reduce the effects of X-ray radiation damage to the protein sample (65). Automated sample injection and data collection were controlled using the *BECQUEREL* beam line control software (66). The SAXS intensities were measured as 2400× 1-s individual X-ray exposures, from the continuously flowing column eluent, using a Pilatus 6M 2D-area detector. The 2D-to-1D data reduction, *i.e.* radial averaging of the data to produce 1D $I(s)$ *versus* s profiles, were performed using the *SASFLOW* pipeline incorporating *RADAVER* from the *ATSAS* 2.8 suite of software tools (67, 68). The 2400 individual frames obtained for each gel filtration-SAXS run were processed using *CHROMIXS* (69). Only those scaled individual SAXS data frames with a consistent R_g through the SEC elution peak after the subtraction of an appropriate solvent blank [as evaluated from the Guinier approximation; $\ln I(s)$ *versus* s^2 for $sR_g < 1.3$] were used to generate the final averaged SAXS profiles. All SAXS data comparisons and data-model fits were assessed using the reduced χ^2 test and the correlation map, or CorMap (33), with p value set at significance threshold of $\alpha = 0.01$. The ambiguity of each data set, with respect to *ab initio* shape restoration, was assessed using *AMBIMETER* (39), whereas the maximum useful data range (s_{max}) and number of Shannon channels was assessed using SHANUM (70). The indirect inverse Fourier transformations of the resulting data for each construct to generate the probable frequency of scattering-pair real-space distances [$p(r)$ profiles] were calculated using *GNOM* (71). The calculation of $p(r)$ and all subsequent modeling was performed within the SHANUM s_{max}

limit. *Ab initio* modeling was performed using *GASBOR* (40) and *DAMMIN (32)*, while atomistic rigid-body modeling, incorporating the crystal structure of the N5 and, where relevant, the *I-TASSER* (34)-generated homology models of N6, N4, and N3 were performed using *BUNCH* (36) and *CORAL* (41), whereas *EOM* (37, 38) was used for ensemble analysis. For the *ab initio* modeling of N6-N3, and because of the ambiguity of the SAXS data, *DAMMIN* and *GASBOR* modeling was performed multiple times (at least 16–30), and from the individual model cohorts, representative clusters of related spatially aligned low-resolution structures were calculated using *DAMCLUST* (41).

All data and models are made available in the Small-Angle Scattering Biological Data Bank (SASBDB) under the accession codes SASDEY9 (N5-N4), SASDEZ9 (N6-N3), and SASDJJ4 (N2). Structural parameter reporting tables (31) for N5-N4 and N6-N3 are reported in supporting information.

Negative-stain EM

For negative-stain EM, affinity-purified N6-N3 (0.17 mg/ml) was freshly diluted 1:100 with TBS. Samples were applied to freshly glow-discharged holey carbon grids with an additional 2-nm thin continuous carbon layer (Quantifoil R2/1 + 2 nm, 200-mesh copper) and stained using uranyl formate (0.75%, w/v). For this, 4.5 µl of the diluted sample was incubated on the grid and then blotted to a thin layer by hand. To remove unbound protein, the grid was washed three times with 15 µl TBS, immediately blotting to a thin liquid layer by hand. The protein was then stained by brief incubation in a droplet of uranyl formate solution, immediate blotting, and incubation in a fresh droplet of uranyl formate for 20 s. Excess liquid was removed by blotting, and the remaining thin layer of stain was rapidly dried by a laminar flow of air.

Transmission electron microscopic images were acquired on a JEOL JEM-2200FS microscope operating at 200 kV, equipped with a 4k × 4k TemCam F-416 camera. Alignment and astigmatism correction were performed manually using the microscope's user interface, and then images were acquired using low-dose condition settings in SerialEM (72). In total, 146 images were collected at a nominal magnification of 50,000×, corresponding to a calibrated pixel size of 2.05 Å pixel^{-1} at a defocus range from 0.02 μm to 3.00 μm.

Data were processed using the Sphire package (73). After defocus estimation and CTF-based quality control, assuming an amplitude contrast of 80%, 146 images were processed by the Phosaurus-network general model using a threshold of 0.1 and a particle size of 140 pixels, corresponding to an object size of ~290 Å, yielding 43,246 particle coordinates. Whereas the choice of a nonrestrictive threshold yielded many false-positive coordinates, as judged by visual inspection, it ensured that all N6-N3 particles were selected despite apparent shape heterogeneity. Particle images were extracted with a box size of 208 pixels and classified at a down-sampled pixel size of 4.95 Å using the ISAC algorithm, specifying 50 particles per class. For visualization, particle images were recomputed at full pixel size with full CTF correction using the beautifier algorithm. In parallel, particle coordinates were imported into the Relion package and processed independently (74). After defocus estimation and CTF-based quality control, assuming an amplitude contrast of 80%, 46,214 particle images were extracted based on the imported coordinates and classified into 100 classes.

Data availability

X-ray data for N2 have been deposited in the Protein Data Bank in Europe with accession code 6SNK. SAXS data and models for N5-N4, N6-N3, and N2 have been deposited in the Small-Angle Scattering Biological Data Bank (SASBDB) under the accession codes SASDEY9, SASDEZ9, and SASDJJ4, respectively. All other data are contained within the manuscript and the supporting information.

Acknowledgments—We thank the staff at the beamline ID30B at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for their support during X-ray diffraction data collection. We thank the staff at the Electron Microscopy and Analytics unit at the research center caesar for their support during EM data collection.

Author contributions—H. S.-D., J. M. G., C. M. J., C. D. F., and P. M. data curation; H. S.-D., J. M. G., C. M. J., U. B., D. I. S., J. M. W., E. B., M. P., and R. W. formal analysis; H. S.-D., J. M. G., C. M. J., E. B., M. P., and R. W. validation; H. S.-D., J. M. G., C. M. J., C. D. F., P. M., L. E. B., and R. J. O. investigation; H. S.-D., J. M. G., C. M. J., C. D. F., and P. M. visualization; H. S.-D., J. M. G., C. M. J., L. E. B., R. J. O., M. P., and R. W. methodology; H. S.-D., J. M. G., and R. W. writing-original draft; H. S.-D., J. M. G., C. M. J., E. B., M. P., and R. W. writing-review and editing; J. M. G., C. M. J., D. I. S., and E. B. software; C. M. J., J. M. W., M. P., and R. W. conceptualization; C. M. J., L. E. B., U. B., D. I. S., R. J. O., E. B., M. P., and R. W. resources; C. M. J., L. E. B., U. B., D. I. S., R. J. O., J. M. W., E. B., M. P., and R. W. supervision; C. M. J., D. I. S., R. J. O., M. P., and R. W. funding acquisition; M. P. and R. W. project administration.

Funding and additional information—This study was supported by the German Research Foundation Grants SFB 829-B2 and FOR 2722-B1 (to M. P. and R. W.) and Grant SFB 829-B11 (to U. B.), and by a UK Medical Research Council Grant MR/K018779/1 (to L. E. B. and R. J. O). Portions of this work was also supported by iNEXT Grant 653706 (to C. M. J. and D. I. S.), funded by the Horizon 2020 program of the European Union (PID: 2861) using the Structural Audit program and an iNEXT proposal ID 5602 (to H. S. D. and R. W.). Crystals were grown in the Cologne Crystallization Facility (C2f.uni-koeln.de; grant no. INST 216/682-1 FUGG from the German Research Foundation).

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: UCMD, Ullrich congenital muscular dystrophy; RMSD, root mean square deviation; ECM, extracellular matrix; BM, Bethlem myopathy; VWA, von Willebrand factor type A; SAXS, small-angle X-ray scattering; SEC-SAXS, size exclusion chromatography SAXS; MALLS, multi-angle laser light scattering; R_g, radius of gyration; D_{max}, maximum

particle dimension; $p(r)$, probable real-space atom-pair distance distribution; RI, refractive index.

References

1. Engel, J., Furthmayr, H., Odermatt, E., von der Mark, H., Aumailley, M., Fleischmajer, R., and Timpl, R. (1985) Structure and macromolecular organization of type VI collagen. *Ann. N. Y. Acad. Sci.* **460,** 25–37 CrossRef Medline
2. Keene, D. R., Engvall, E., and Glanville, R. W. (1988) Ultrastructure of type VI collagen in human skin and cartilage suggests an anchoring function for this filamentous network. *J. Cell Biol.* **107,** 1995–2006 CrossRef Medline
3. Kuo, H. J., Maslen, C. L., Keene, D. R., and Glanville, R. W. (1997) Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. *J. Biol. Chem.* **272,** 26522–26529 CrossRef Medline
4. Groulx, J. F., Gagné, D., Benoit, Y. D., Martel, D., Basora, N., and Beaulieu, J. F. (2011) Collagen VI is a basement membrane component that regulates epithelial cell-fibronectin interactions. *Matrix Biol.* **30,** 195–206 CrossRef Medline
5. Wiberg, C., Hedbom, E., Khairullina, A., Lamande, S. R., Oldberg, A., Timpl, R., Morgelin, M., and Heinegard, D. (2001) Biglycan and decorin bind close to the N-terminal region of the collagen VI triple helix. *J. Biol. Chem.* **276,** 18947–18952 CrossRef Medline
6. Cescon, M., Gattazzo, F., Chen, P., and Bonaldo, P. (2015) Collagen VI at a glance. *J. Cell Sci.* **128,** 3525–3531 CrossRef Medline
7. Lampe, A. K., and Bushby, K. M. (2005) Collagen VI related muscle disorders. *J. Med. Genet.* **42,** 673–685 CrossRef Medline
8. Jobsis, G. J., Bolhuis, P. A., Boers, J. M., Baas, F., Wolterman, R. A., Hensels, G. W., and De Visser, M. (1996) Genetic localization of Bethlem myopathy. *Neurology* **46,** 779–782 CrossRef Medline
9. Jobsis, G. J., Boers, J. M., Barth, P. G., and de Visser, M. (1999) Bethlem myopathy: a slowly progressive congenital muscular dystrophy with contractures. *Brain* **122,** 649–655 CrossRef
10. Higuchi, I., Shiraishi, T., Hashiguchi, T., Suehara, M., Niiyama, T., Nakagawa, M., Arimura, K., Maruyama, I., and Osame, M. (2001) Frameshift mutation in the collagen VI gene causes Ullrich's disease. *Ann. Neurol.* **50,** 261–265 CrossRef Medline
11. Demir, E., Sabatelli, P., Allamand, V., Ferreiro, A., Moghadaszadeh, B., Makrelouf, M., Topaloglu, H., Echenne, B., Merlini, L., and Guicheney, P. (2002) Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. *Am. J. Hum. Genet.* **70,** 1446–1458 CrossRef Medline
12. Gara, S. K., Grumati, P., Urciuolo, A., Bonaldo, P., Kobbe, B., Koch, M., Paulsson, M., and Wagener, R. (2008) Three novel collagen VI chains with high homology to the alpha3 chain. *J. Biol. Chem.* **283,** 10658–10670 CrossRef Medline
13. Fitzgerald, J., Rich, C., Zhou, F. H., and Hansen, U. (2008) Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). *J. Biol. Chem.* **283,** 20170–20180 CrossRef Medline
14. Schmid, T. M., and Linsenmayer, T. F. (1983) A short chain (pro)collagen from aged endochondral chondrocytes. Biochemical characterization. *J. Biol. Chem.* **258,** 9504–9509
15. Jander, R., Rauterberg, J., and Glanville, R. W. (1983) Further characterization of the three polypeptide chains of bovine and human short-chain collagen (intima collagen). *Eur. J. Biochem.* **133,** 39–46 CrossRef Medline
16. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250,** 194–199 CrossRef Medline
17. Odermatt, E., Risteli, J., van Delden, V., and Timpl, R. (1983) Structural diversity and domain composition of a unique collagenous fragment (intima collagen) obtained from human placenta. *Biochem. J.* **211,** 295–302 CrossRef Medline
18. Knupp, C., and Squire, J. M. (2001) A new twist in the collagen story–the type VI segmented supercoil. *EMBO J.* **20,** 372–376 CrossRef Medline
19. Heumuller, S. E., Talantikite, M., Napoli, M., Armengaud, J., Morgelin, M., Hartmann, U., Sengle, G., Paulsson, M., Moali, C., and Wagener, R. (2019) C-terminal proteolysis of the collagen VI alpha3 chain by BMP-1 and proprotein convertase(s) releases endotrophin in fragments of different sizes. *J. Biol. Chem.* **294,** 13769–13780 CrossRef Medline
20. Beecher, N., Roseman, A. M., Jowitt, T. A., Berry, R., Troilo, H., Kammerer, R. A., Shuttleworth, C. A., Kielty, C. M., and Baldock, C. (2011) Collagen VI, conformation of A-domain arrays and microfibril architecture. *J. Biol. Chem.* **286,** 40266–40275 CrossRef Medline
21. Godwin, A. R., Starborg, T., Sherratt, M. J., Roseman, A. M., and Baldock, C. (2017) Defining the hierarchical organisation of collagen VI microfibrils at nanometre to micrometre length scales. *Acta Biomater.* **52,** 21–32 CrossRef Medline
22. Maaë, T., Bayley, C. P., Mörgelin, M., Lettmann, S., Bonaldo, P., Paulsson, M., Baldock, C., and Wagener, R. (2016) Heterogeneity of collagen VI microfibrils: structural analysis of non-collagenous regions. *J. Biol. Chem.* **291,** 5247–5258 CrossRef Medline
23. Becker, A. K., Mikolajek, H., Paulsson, M., Wagener, R., and Werner, J. M. (2014) A structure of a collagen VI VWA domain displays N and C termini at opposite sides of the protein. *Structure* **22,** 199–208 CrossRef Medline
24. Butterfield, R. J., Foley, A. R., Dastgir, J., Asman, S., Dunn, D. M., Zou, Y., Hu, Y., Donkervoort, S., Flanigan, K. M., Swoboda, K. J., Winder, T. L., Weiss, R. B., and Bönnemann, C. G. (2013) Position of glycine substitutions in the triple helix of COL6A1, COL6A2, and COL6A3 is correlated with severity and mode of inheritance in collagen VI myopathies. *Hum. Mutat.* **34,** 1558–1567 CrossRef Medline
25. Pan, T. C., Zhang, R. Z., Pericak-Vance, M. A., Tandan, R., Fries, T., Stajich, J. M., Viles, K., Vance, J. M., Chu, M. L., and Speer, M. C. (1998) Missense mutation in a von Willebrand factor type A domain of the alpha 3 (VI) collagen gene (COL6A3) in a family with Bethlem myopathy. *Hum. Mol. Genet.* **7,** 807–812 CrossRef Medline
26. Baker, N. L., Mörgelin, M., Pace, R. A., Peat, R. A., Adams, N. E., Gardner, R. J. M., Rowland, L. P., Miller, G., De Jonghe, P., Ceulemans, B., Hannibal, M. C., Edwards, M., Thompson, E. M., Jacobson, R., Quinlivan, R. C. M., *et al.* (2007) Molecular consequences of dominant Bethlem myopathy collagen VI mutations. *Ann. Neurol.* **62,** 390–405 CrossRef Medline
27. Lampe, A. K., Dunn, D. M., von Niederhausern, A. C., Hamil, C., Aoyagi, A., Laval, S. H., Marie, S. K., Chu, M.-L., Swoboda, K., Muntoni, F., Bonnemann, C. G., Flanigan, K. M., Bushby, K. M. D., and Weiss, R. B. (2005) Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J. Med. Genet.* **42,** 108–120 CrossRef Medline
28. Marakhonov, A. V., Tabakov, V. Y., Zernov, N. V., Dadali, E. L., Sharkova, I. V., and Skoblov, M. Y. (2018) Two novel COL6A3 mutations disrupt extracellular matrix formation and lead to myopathy from Ullrich congenital muscular dystrophy and Bethlem myopathy spectrum. *Gene* **672,** 165–171 CrossRef Medline
29. Sasaki, T., Hohenester, E., Zhang, R. Z., Gotta, S., Speer, M. C., Tandan, R., Timpl, R., and Chu, M. L. (2000) A Bethlem myopathy Gly to Glu mutation in the von Willebrand factor A domain N2 of the collagen alpha3(VI) chain interferes with protein folding. *FASEB J.* **14,** 761–768 CrossRef Medline
30. Whittaker, C. A., and Hynes, R. O. (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol. Biol. Cell* **13,** 3369–3387 CrossRef Medline
31. Graewert, M. A., Franke, D., Jeffries, C. M., Blanchet, C. E., Ruskule, D., Kuhle, K., Flieger, A., Schafer, B., Tartsch, B., Meijers, R., and Svergun, D. I. (2015) Automated pipeline for purification, biophysical and X-ray analysis of biomacromolecular solutions. *Sci. Rep.* **5,** 10734 CrossRef Medline
32. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 CrossRef
33. Franke, D., Jeffries, C. M., and Svergun, D. I. (2015) Correlation map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12,** 419–422 CrossRef Medline
34. Yang, J. Y., Yan, R. X., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 CrossRef Medline
35. Suhre, M. H., Gertz, M., Steegborn, C., and Scheibel, T. (2014) Structural and functional features of a collagen-binding matrix protein from the mussel byssus. *Nat. Commun.* **5,** 3392 CrossRef Medline

36. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89,** 1237–1250 CrossRef Medline
37. Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2,** 207–217 CrossRef Medline
38. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129,** 5656–5664 CrossRef Medline
39. Petoukhov, M. V., and Svergun, D. I. (2015) Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr. D Biol. Crystallogr.* **71,** 1051–1058 CrossRef Medline
40. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80,** 2946–2953 CrossRef Medline
41. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 CrossRef Medline
42. Baldock, C., Sherratt, M. J., Shuttleworth, C. A. A., and Kielty, C. M. (2003) The supramolecular organization of collagen VI microfibrils. *J. Mol. Biol.* **330,** 297–307 CrossRef Medline
43. Fitzgerald, J., Morgelin, M., Selan, C., Wiberg, C., Keene, D. R., Lamande, S. R., and Bateman, J. F. (2001) The N-terminal N5 subdomain of the alpha 3(VI) chain is important for collagen VI microfibril formation. *J. Biol. Chem.* **276,** 187–193 CrossRef Medline
44. Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J., and Chu, M. L. (1992) Structure of recombinant N-terminal globule of type VI collagen alpha 3 chain and its binding to heparin and hyaluronan. *EMBO J.* **11,** 4281–4290 CrossRef Medline
45. Lamandé, S. R., and Bateman, J. F. (2018) Collagen VI disorders: Insights on form and function in the extracellular matrix and beyond. *Matrix Biol.* **71-72,** 348–367 CrossRef Medline
46. Becker, A. K., Mikolajek, H., Werner, J. M., Paulsson, M., and Wagener, R. (2015) Characterization of recombinantly expressed matrilin VWA domains. *Protein Expr. Purif.* **107,** 20–28 CrossRef Medline
47. Camacho Vanegas, O., Bertini, E., Zhang, R. Z., Petrini, S., Minosse, C., Sabatelli, P., Giusti, B., Chu, M. L., and Pepe, G. (2001) Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc. Natl. Acad. Sci. U S A* **98,** 7516–7521 CrossRef Medline
48. Lampe, A. K., Zou, Y., Sudano, D., O'Brien, K. K., Hicks, D., Laval, S. H., Charlton, R., Jimenez-Mallebrera, C., Zhang, R.-Z., Finkel, R. S., Tennekoon, G., Schreiber, G., van der Knaap, M. S., Marks, H., Straub, V., *et al.* (2008) Exon skipping mutations in collagen VI are common and are predictive for severity and inheritance. *Hum. Mutat.* **29,** 809–822 CrossRef Medline
49. Foley, A. R., Hu, Y., Zou, Y., Yang, M., Medne, L., Leach, M., Conlin, L. K., Spinner, N., Shaikh, T. H., Falk, M., Neumeyer, A. M., Bliss, L., Tseng, B. S., Winder, T. L., and Bönnemann, C. G. (2011) Large genomic deletions: a novel cause of Ullrich congenital muscular dystrophy. *Ann. Neurol.* **69,** 206–211 CrossRef Medline
50. Zhang, R. Z., Zou, Y., Pan, T. C., Markova, D., Fertala, A., Hu, Y., Squarzoni, S., Reed, U. C., Marie, S. K., Bonnemann, C. G., and Chu, M. L. (2010) Recessive COL6A2 C-globular missense mutations in Ullrich congenital muscular dystrophy: role of the C2a splice variant. *J. Biol. Chem.* **285,** 10005–10015 CrossRef Medline
51. Chen, J.-M., Férec, C., and Cooper, D. N. (2009) Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. *Hum. Mutat.* **30,** 1435–1448 CrossRef
52. Tuckwell, D. (1999) Evolution of von Willebrand factor A (VWA) domains. *Biochem. Soc. Trans.* **27,** 835–840 CrossRef Medline
53. Mazzucato, M., Spessotto, P., Masotti, A., De Appollonia, L., Cozzi, M. R., Yoshioka, A., Perris, R., Colombatti, A., and De Marco, L. (1999) Identification of domains responsible for von Willebrand factor type VI collagen interaction mediating platelet adhesion under high flow. *J. Biol. Chem.* **274,** 3033–3041 CrossRef Medline
54. Baldauf, C., Schneppenheim, R., Stacklies, W., Obser, T., Pieconka, A., Schneppenheim, S., Budde, U., Zhou, J., and Grater, F. (2009) Shear-induced unfolding activates von Willebrand factor A2 domain for proteolysis. *J. Thromb. Haemost.* **7,** 2096–2105 CrossRef Medline
55. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74,** 85–97 CrossRef Medline
56. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 282–292 CrossRef Medline
57. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 CrossRef Medline
58. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 CrossRef Medline
59. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of COOT. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 CrossRef Medline
60. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 CrossRef Medline
61. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 CrossRef Medline
62. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., and Svergun, D. I. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48,** 431–443 CrossRef Medline
63. Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E., and Svergun, D. I. (2016) Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **11,** 2122–2153 CrossRef Medline
64. Brown, P. H., and Schuck, P. (2008) A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput. Phys. Commun.* **178,** 105–120 CrossRef Medline
65. Jeffries, C. M., Graewert, M. A., Svergun, D. I., and Blanchet, C. E. (2015) Limiting radiation damage for high-brilliance biological solution scattering: practical experience at the EMBL P12 beamline PETRAIII. *J. Synchrotron Radiat.* **22,** 273–279 CrossRef Medline
66. Hajizadeh, N. R., Franke, D., and Svergun, D. I. (2018) Integrated beamline control and data acquisition for small-angle X-ray scattering at the P12 BioSAXS beamline at PETRAIII storage ring DESY. *J. Synchrotron Radiat.* **25,** 906–914 CrossRef Medline
67. Franke, D., Kikhney, A. G., and Svergun, D. I. (2012) Automated acquisition and analysis of small angle X-ray scattering data. *Nucl. Instrum. Methods Phys. Res.* **689,** 52–59 CrossRef
68. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 CrossRef Medline
69. Panjkovich, A., and Svergun, D. I. (2018) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34,** 1944–1946 CrossRef Medline
70. Konarev, P. V., and Svergun, D. I. (2015) A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ* **2,** 352–360 CrossRef Medline

71. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25,** 495–503 CrossRef
72. Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152,** 36–51 CrossRef Medline
73. Moriya, T., Saur, M., Stabrin, M., Merino, F., Voicu, H., Huang, Z., Penczek, P. A., Raunser, S., and Gatsogiannis, C. (2017) High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.* CrossRef
74. Scheres, S. H. W. (2012) A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415,** 406–418 CrossRef Medline