A new twist in the collagen story—the type VI segmented supercoil

Carlo Knupp and John M.Squire¹

Biological Structure and Function Section, Biomedical Sciences Division, Imperial College of Science, Technology & Medicine, London SW7 2AZ, UK

¹Corresponding author
e-mail: j.squire@ic.ac.uk

Collagen occurs in two major forms: fibrillar and non-fibrillar. Non-fibrillar collagens are structurally more variable and relatively ill-understood. In this work we analysed the amino acid sequence of type VI collagen, a non-fibrillar collagen that forms antiparallel dimers. A sequence motif was discovered that gives rise to systematic molecular coiling. There is a common periodicity (~23 or 2×23 residues) in the charged amino acids, in the prolines and in the discontinuities in the Gly-X-Y triplets. In addition, there is a different periodicity (~21 amino acids) in the apolar groups. The two repeats mean that the only way to simultaneously maximize both the hydrophobic and polar interactions during dimer formation is with the molecules antiparallel, overlapped by 75 nm as observed, and supercoiled. The alternating proline-rich and charge-rich patches, often together with discontinuities in the Gly-X-Y sequences, coincide with each half-turn of the supercoil, thus breaking it into segments. We have termed this structure the collagen segmented supercoil. The segmented supercoil and variants may be common aggregation motifs for the non-fibrillar collagens.

Keywords: amino acid sequence analysis/collagen IV/collagen VI/collagen segmented supercoil/ non-fibrillar collagen

Introduction

Unlike the α -proteins, where Francis Crick's coiled coil is a well-known structural motif (Crick, 1952; Lupas, 1996), to date there has been little evidence for an analogous supercoiled structure between pairs of collagen molecules. In this paper we report findings from the analysis of the type VI collagen amino acid sequence that suggest the presence of analogous supercoiled structural motifs in collagen as well.

Type VI collagen anchors large interstitial structures, such as nerves, blood vessels and collagen fibrils, into surrounding connective tissue. It is also involved in cell migration and differentiation, and may play a role in bridging cells with the extracellular matrix (Ricard-Blum *et al.*, 2000). Genetic mutations of collagen VI have been linked to Bethlem myopathy, an inherited disorder characterized by early childhood onset of generalized

muscle weakness and wasting and, commonly, contracture of multiple joints (Lamande *et al.*, 1999).

Type VI collagen molecules (Furthmayr *et al.*, 1983) comprise an extended central rod-like triple-helical region ~105 nm long, on each end of which are globular domains (Figure 1A). The type VI molecule is a heterotrimer of three genetically distinct $\alpha 1(VI)$, $\alpha 2(VI)$ and $\alpha 3(VI)$ chains (Jander *et al.*, 1983). The three α -chains of type VI collagen contain a short collagenous sequence of ~335/336 amino acids each, which are flanked at the N- and C-termini by non-collagenous sequences predicted to form large globular domains. The non-collagenous sequences are dominated by numerous A-domains (Shuttleworth *et al.*, 1999). Molecules are assembled intracellularly, first into antiparallel dimers, then into tetramers, and sometimes into larger aggregates (beaded filaments; Furthmayr *et al.*, 1983; Ayad *et al.*, 1994; Knupp *et al.*, 2000). Dimers and tetramers are stabilized by inter- and intramolecular disulfide bonds (Ayad *et al.*, 1994). Studies are presently being carried out to understand the role of the A-domains during trimerization and higher-order intracellular oligomerization (Shuttleworth *et al.*, 1999). The regular Gly-X-Y repeat in the type VI triple-helical region is occasionally broken by discontinuities, in some ways analogous to skips (stammers and stutters) in the heptad sequences of coiled-coil α -proteins (Brown *et al.*, 1996). Observations of isolated dimers of type VI collagen under the electron microscope (Furthmayr *et al.*, 1983) have concluded that type VI molecules pack in an antiparallel fashion with a molecular overlap of ~75 nm, leaving non-overlapped single molecular ends of ~30 nm (Figure 1A). The antiparallel molecules are thought to be linked by disulfide bridges at each end of the 75 nm central segment (Furthmayr *et al.*, 1983; Chu *et al.*, 1988). Some electron micrographs have shown hints of molecular twisting in the central domain of the antiparallel dimer (Furthmayr *et al.*, 1983) with possibly four or five cross-overs within the 75 nm domain length. The present study asks first what stabilizes the interactions between the two antiparallel molecules and secondly whether there is anything in the type VI sequence that suggests why the molecules might supercoil.

Results

Fourier transform computations of the summed polar scores for the three chains (Figure 1B) gave a clear periodicity of ~46.5 amino acids (range 48.7–44.5), with a weaker peak at 36.5 (37.9–35.3). Similar analysis of the proline residues (Figure 1C) gave a repeat of 23.3 amino acids (range 23.8–22.7). The discontinuities in the tripeptide sequences (e.g. the absence of glycine at the expected position or a phase change in the Gly-X-Y sequence) occurred at approximate intervals of 44–48

amino acids, apart from one missing discontinuity. Clearly, all these repeats are related. On the other hand, the hydrophobic sequence (Figure 1D) gave a clear repeat of 21.3 residues (range 20.9–21.8). Inverse Fourier transformation from these transform peaks, showing an averaged (sinusoidal) distribution of whatever feature is being analysed, are very revealing (Figure 2A). Taking the first two 23.3 amino acid proline repeats at the N-terminus as a starting point, the negative charges dominate in the first of these repeats (proximal to the N-terminus) and the positive charges dominate in the second. This pattern then repeats along the molecule. The effect of including the weaker polar repeat at 36.5 amino acids in the Fourier synthesis is to modulate the distribution of charge slightly without altering the main features of the charge profile in Figure 2A. The sequence discontinuities occur regularly in the middle of every second proline repeat. The hydrophobic repeat is such that it starts at the N-terminal end (left-hand edge of Figure 2A), more or less in phase with the proline repeat. Moving to the right, the hydrophobics then become gradually out of step with the prolines so that they are in antiphase halfway along the molecule and are back exactly in phase at the right-hand (C-terminal) end. This analysis (Figure 2A) shows that the hydrophobic pattern and the proline pattern have approximate mirror symmetry about the centre of this sequence, whereas the distribution of charges is asymmetric and complementary.

Taking the amino acid distribution of one molecule and trying to match against it the appropriate features of the adjacent molecule in the pair reveals the following properties. (i) If the interacting molecules are parallel, then they must be axially shifted to bring the positive charges on one molecule opposite the negative charges of the other. However, if this is done, the hydrophobic residues, which have a different repeat, must then inevitably be out of step in the two molecules and the cystine residues will also not line up. (ii) If the interacting molecules are antiparallel, then they could either be overlapped by 75 nm, as in Figure 2D, or they could be stepped in either direction by a small multiple of the 48 polar amino acid repeat in order to optimize the interactions of opposite charges. (iii) However, in (ii), if the molecules are overlapped by anything other than 75 nm, then the hydrophobic patches are no longer in optimal register. The only structure that brings opposite charges into alignment and also optimizes apolar interactions at the same time is the 75 nm overlapped antiparallel dimer. (iv) With this 75 nm overlap, the cystine residues at the ends of the tripeptide regions of the two chains are brought into axial alignment, where, through disulfide bridge formation, they can stabilize the structure generated by the apolar and charged residue interactions.

At this point, there are good reasons to see why the type VI molecules in a dimer are antiparallel and overlapped by 75 nm. However, in order to investigate the possible presence of supercoiling, the structure needs to be probed in more detail. First, analysis of the separate chains shows that the $\alpha 1$ chain contains a higher than average number of negative charges (chick net charge -11 [$+25$, -36]), that the $\alpha 3$ chain contains a higher than average number of positive charges (chick net charge $+20$ [$+42$, -22]) and that the $\alpha 2$ chain is close to neutral (chick net charge $+3$ [$+36$, -33]). When analysing the hydro-

Fig. 1. Schematic diagrams of collagen VI dimers and the sequence Fourier transforms from the overlap region. (A) Two 105-nm-long molecules assemble antiparallel with a 75 nm overlap and are linked by disulfide bridges. Electron microscopic evidence (Furthmayr *et al.*, 1983) hints at molecular supercoiling in the 75-nm-long central region of the dimer (lower figure). (B) Linear Fourier transform of the summed polar amino acid scores for the three α -chains, showing the main peak (white arrowhead) at 46.5 amino acids used for the Fourier synthesis (blue and red lines) in Figure 2A. Sequences used were summed to give a 'property' (e.g. charge) sequence length of ~250 and then embedded in a matrix of zeros to give an array length of 1024 for Fourier transformation. Incorporation of the second peak at ~36.5 modulates the Fourier synthesis profile in Figure 2A, but does not change its general features. (C) Linear Fourier transform of the summed proline scores for the three α -chains [method as in (B)]. The main peak at 23.3 amino acids (white arrowhead) was used in the Fourier synthesis (white line) in Figure 2A. (D) Linear Fourier transform of the hydrophobic amino acids along the sum of the three α -chains [method as in (B)], with a main peak at 21.3 amino acids used for the Fourier synthesis (yellow line) in Figure 2A.

phobic residues in individual chains (Figure 2B), we found that the main contribution to the total molecular hydrophobicity towards the N-terminal end is from the $\alpha 3$ chain, that at the C-terminal end is from the $\alpha 1$ chain and that in the middle is from $\alpha 2$, with slightly overlapping contributions from $\alpha 1$ and $\alpha 3$. If the amino acid characteristics had been evenly spread along all three chains, then there would have been no evidence for supercoiling. However, one would imagine that here, as in the coiled-coil α -helical structures, the main stabilizing feature on the line of contact of the two molecules will be the apolar interactions. The fact that the sites of greatest hydrophobicity follow roughly helical tracks along the molecules (Figure 2C) and shift from one chain to another suggests that supercoiling must occur. The $\alpha 123$ molecule shown in Figure 2C would give a supercoil of pitch ~ 37.5 nm (roughly one-half of 75 nm). Note that the supercoil has the opposite sense of twist (left-handed) to the underlying twist of the individual collagen chains (right-handed), which in turn is opposite to the twist within each chain (left-handed), just as has been found in all other supercoiled molecules (e.g. the α -helical coiled coil). This is the characteristic of 'regular lay' rope structures and confers strength and rigidity on the assembly.

Finally, since the sequence discontinuities (when they occur) are precisely in step with the proline and polar amino acid repeats, one can imagine a pseudo repeating unit comprising: (break) – short proline region – negative patch – long proline region – positive patch – short proline region – (break). However, the hydrophobic repeat does not coincide with this main structural sequence, so that in terms of their apolar character every one of these repeats is different. The presence of the repeats separated by discontinuities gives the supercoil a clearly segmented character, hence our preference for the name 'segmented supercoil' for this collagen structure. This segmentation is also emphasized by the fact that regions of high proline content are likely to be rather rigid, whereas the polar-rich

regions are likely to be more flexible. We would, therefore, not expect the curvature along the supercoil to be constant, but to vary slightly within each segment.

Conclusion

In conclusion, it is evident from the mixed periodicities of the hydrophobic and polar amino acids in type VI collagen molecules why antiparallel dimerization occurs, why the observed 75 nm overlap is preferred and why there is supercoiling. Type VI collagen is an example of a non-fibrillar collagen. The much longer type IV collagen found in the basal lamina is another member of this family. Since type VI collagen shares some common features with parts of the type IV molecule (Yurchenco and Schittny, 1990), and there is also microscopic evidence for aggregation and molecular twisting (Yurchenco and Ruben, 1987), it seems probable that there may be further examples of antiparallel dimer aggregation, possibly involving type IV collagen and some of the other collagen types. Like the α -helix coiled coil, the collagen segmented supercoil may be a common motif.

Materials and methods

Sequence analysis of the central 75 nm domain was carried out by analysing separately the contributions of the positively and negatively charged residues (1), the hydrophobic residues (2), and also the numerous prolines in positions X and Y (3). The sequences of the three α -chains [from Swissprot CA16_CHICK (accession Nos P20785, residues 342–590), CA26_CHICK (P15988, 342–590) and CA36_CHICK (P15989, 2130–2378)] were projected into three corresponding unidimensional arrays, observing the following transformation rules: (1) charged residues were substituted as 1 (H,K,R) or –1 (D,E) and all other residues as 0; (2) hydrophobic residues were substituted by their absolute value in the hydrophobicity scale of Kyte and Doolittle (1982), while all other amino acids were projected as 0; (3) prolines were substituted as 1 and all other amino acids as 0. In all cases, these arrays were either used for each chain individually or were summed over the three chains for further analysis. The total amino acid array length was ~ 250 residues and this was extended by zeros to make a 1024 array for Fourier transformation. Also, the sequence scores were smoothed locally

Fig. 2. Sequence character along the type VI collagen overlap region and illustration of the segmented coil model of dimer formation. **(A)** Fourier syntheses of the polar, hydrophobic and proline residue distributions in the 75-nm-long region of the collagen VI dimer using the principal peaks in Figure 1B–D. Blue, the charge distribution along a molecule; red, the same charge distribution reversed as for the second (antiparallel) molecule; white, the proline distribution of each molecule; yellow, the hydrophobic distribution of each molecule. Since the proline and hydrophobic distributions are symmetrical about the middle, the illustrated distributions apply to both members of the antiparallel pair. On the other hand, the distribution of polar residues is asymmetric and complementary across the midline. This means that areas of high positive charge density along one molecule (blue line) are in register with minima (areas of high negative charge density) of the other molecule (red line). Charge/proline distributions and hydrophobic distributions are in phase towards the extremities of the 75-nm-long region, but out of phase in the central part. Green arrows indicate the positions of the interruptions along the two portions of the triple helix. These generally occur in the middle of proline-rich regions. **(B)** Hydrophobicity profiles in the 75-nm-long region of the type VI dimers. In the three panels, the abscissa represents the amino acid position along the triple helix, while the ordinate is the hydrophobicity score. The curve underlying the light blue areas is the sum of the hydrophobic contribution of all three α -chains. In dark blue, red and yellow are highlighted the individual contributions from the $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains, respectively. The hydrophobic contributions are not evenly spread along the three chains; the main contributions are from $\alpha 3$ towards the N-terminal end (left), from $\alpha 1$ towards the C-terminal end and from $\alpha 2$ in the middle, with slightly overlapping contributions here from $\alpha 1$ and $\alpha 3$. These regions are highlighted by yellow, blue and red background shading, respectively. Black arrows highlight the 21–22 amino acid repeat of the total hydrophobic (light blue) distribution. **(C)** Radial projection of the collagen triple helix in the 75-nm-long region of type VI collagen that contributes to dimerization. This radial projection can be thought of as being produced by wrapping a piece of paper into a cylinder around the collagen triple helix and then tracing the positions of the single α -chains onto it; the $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains are shown in blue, red and yellow, respectively. The black circles represent the locations of the main hydrophobic regions on the triple helix (their repeat is indicated by the black arrowheads). These amino acids lie on a track corresponding to a left-handed superhelix of pitch 37.5 nm. Two antiparallel molecules (one represented in black and the other in green) can only bring together the hydrophobic patches along the corresponding two superhelices if the molecules twist around each other. **(D)** Diagram of the 75-nm-long region contributing to dimerization on two antiparallel molecules. Red and blue regions show where there are bands of net positive or negative charge; grey patches show regions of high proline density. Green arrows highlight regions where the triple helix Gly-X-Y sequence is interrupted and yellow arrows indicate the location of the cystines. The positions of the main hydrophobic patches are indicated by ovals around the triple helix (black in front, grey behind). **(E)** Model of the collagen VI segmented supercoil showing the complementary polar interaction of the molecular bands illustrated in (D) (colour code the same). The hydrophobic patches shown in (D) are now buried on the central line of contact of the two molecules. Bar (applies to all parts of the figure) = 50 amino acids or 15 nm.

by an approximate Gaussian with a 3–4 amino acid half-width, reflecting the sphere of influence particularly of the polar groups. Fourier transforms are a powerful mathematical tool used to detect the periodic behaviour of continuous functions. In fact, continuous functions, such as the unidimensional arrays described above, can be represented by a summation of cosine waves with appropriate amplitude, frequency and phase. By Fourier transforming these functions, a graph is created in which the squares of the amplitudes (intensity) of the constituent cosine waves are plotted with respect to their frequency. Functions that possess a periodic behaviour, i.e. with one of the cosine waves being predominant with respect to the others, are easily recognized in Fourier transforms by the presence of major peaks. Amplitudes, frequencies and phases of the cosine waves corresponding to major peaks are readily inferred [see Parry (1979) and Bracewell (1986) for mathematical details]. Swissprot data manipulation was carried out using Microsoft Word and Excel macros. The Fourier transform program used in this analysis was written using Microsoft Visual Basic and the results were displayed with Microsoft Excel. The figures were prepared with Corel Draw.

An uncertainty in the analysis is the precise axial alignment of the three chains in the collagen triple helix. There is an axial shift of one amino acid between the three chains in the 10/3 helix and there is a 'hand' (clockwise or anticlockwise) to the order of the chain starts along the helix. Writing $\alpha 1(VI)$, $\alpha 2(VI)$, $\alpha 3(VI)$ as $\alpha 123$, the six possible permutations of the three chains are in two groups: one group is $\alpha 123$, $\alpha 312$ and $\alpha 231$, and the second is $\alpha 321$, $\alpha 132$ and $\alpha 213$. Without local smoothing, the distinction between the $\alpha 123$ triple helix and its cyclic permutations is necessary since the three chains are axially shifted in the triple helix such that position X in the first chain is opposite position Y in the second and opposite the glycine of the third. With the smoothing, we have, therefore, analysed only a representative of each of the two groups, namely $\alpha 123$ and $\alpha 321$. The results are broadly the same for both groups and we therefore refer only to $\alpha 123$.

Although the analysis presented here was carried out on collagen VI from chick, analogous results are obtainable from the human form. Collagen VI from other species was not analysed since complete amino acid sequences are not publicly available.

Acknowledgements

This work was supported by a programme grant to Charles Michel and J.M.S. from the Wellcome Trust (#038904) and a Wellcome Trust Prize Studentship for C.K. (#050032).

References

- Ayad, S., Boot-Hanford, R.P., Humphries, M.J., Kadler, K.E. and Shuttleworth, A. (1994) *The Extracellular Matrix Facts Book*. Academic Press, London, UK.
- Bracewell, R.N. (1986) *The Fourier Transform and its Applications*. McGraw-Hill, New York, NY.
- Brown, J., Cohen, C. and Parry, D.A.D. (1996) Heptad breaks in α -helical coiled-coils: stutters and stammers. *Proteins*, **26**, 134–145.
- Chu, M.L., Conway, D., Pan, T.C., Baldwin, C., Mann, K., Deutzmann, R. and Timpl, R. (1988) Amino acid sequence of the triple helical domain of human collagen Type VI. *J. Biol. Chem.*, **263**, 18601–18606.
- Crick, F.H.C. (1952) Is α -keratin a coiled-coil? *Nature*, **170**, 882–883.
- Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. and Engel, J. (1983) Electron microscopical approach to a structural model of intima collagen. *Biochem. J.*, **211**, 303–311.
- Jander, R., Rauterberg, J. and Glanville, R.W. (1983) Further characterisation of the three polypeptide chains of bovine and human short chain collagen (intima collagen). *Eur. J. Biochem.*, **133**, 39–46.
- Knupp, C., Munro, P.M.G., Luther, P.K., Ezra, E. and Squire, J.M. (2000) Structure of abnormal molecular assemblies (collagen VI) associated with human full thickness macular holes. *J. Struct. Biol.*, **129**, 38–47.
- Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–142.
- Lamande, S.R., Shields, K.A., Kornberg, A.J., Shield, L.K. and Bateman, J.F. (1999) Bethlem myopathy and engineered collagen VI triple helical deletion prevent intracellular multimer-assembly and protein secretion. *J. Biol. Chem.*, **274**, 21817–21822.
- Lupas, A. (1996) Coiled-coils: new structures and new functions. *Trends Biochem. Sci.*, **21**, 375–382.
- Parry, D.A.D. (1979) Determination of structural information from the amino acid sequences of fibrous proteins. In Parry, D.A.D. and Creamer, L.K. (eds), *Fibrous Proteins: Scientific, Industrial and Medical Aspects*. Vol. 1. Academic Press, London, UK, pp. 393–427.
- Ricard-Blum, S., Dublet, B. and van der Rest, M. (2000) *Unconventional Collagens*. Oxford University Press, Oxford, UK, pp. 14–20.
- Shuttleworth, A., Ball, S., Baldock, C., Fakhoury, H. and Kielty, C.M. (1999) Functional role of A-domains in type VI collagen. *Biochem. Soc. Trans.*, **27**, 821–823.
- Yurchenco, P.D. and Ruben, G.C. (1987) Basement membrane structure *in situ*: evidence for lateral associations in the type IV collagen network. *J. Cell Biol.*, **105**, 2559–2568.
- Yurchenco, P.D. and Schittny, J.C. (1990) Molecular architecture of basement membranes. *FASEB J.*, **4**, 1577–1590.

Received September 6, 2000; revised and accepted December 5, 2000