IN SITU D-PERIODIC MOLECULAR STRUCTURE OF TYPE II COLLAGEN

Olga Antipova\textsuperscript{1,2,*} and Joseph P.R.O. Orgel\textsuperscript{1,2,*}

\textsuperscript{1}BioCAT and \(\mu\)CoSM Centers: Pritzker Institute of Biomedical Science and Engineering, Illinois Institute of Technology, 3440 S. Dearborn Ave, Chicago, IL 60616, USA.

\textsuperscript{2}CSRRI and Department of Biological, Chemical and Physical Sciences, Illinois Institute of Technology, 3101 S. Dearborn Ave, Chicago, IL 60616, USA.

Running head: D-periodic structure of type II collagen

*Corresponding Authors: J.O. orgel@iit.edu; O.A. antiolg@iit.edu

Collagens are essential components of extracellular matrices in multicellular animals. Fibrillar type II collagen is the most prominent component of articular cartilage and other cartilage-like tissues such as notochord. Its \textit{in situ} macromolecular and packing structures have not been fully characterized, but an understanding of these attributes may help reveal mechanisms of tissue assembly and degradation (as in osteo- and rheumatoid arthritis). In some tissues such as lamprey notochord, the collagen fibrillar organization is naturally crystalline and may be studied by X-ray diffraction. We used diffraction data from native and derivative notochord tissue samples to solve the axial, D-periodic structure of type II collagen via Multiple Isomorphous Replacement. The electron density maps and heavy atom data revealed the conformation of the non-helical telopeptides and the overall D-periodic structure of collagen type II in native tissues, data that were further supported by structure prediction and transmission electron microscopy (TEM). These results help to explain the observed differences in collagen type I and type II fibrillar architecture and indicate the collagen type II crosslink organization, which is crucial for fibrillogenesis. TEM data show the close relationship between lamprey and mammalian collagen fibrils, even though the respective larger scale tissue architecture differs.

Type II collagen fibrils are major components of cartilage, intervertebral discs, and the vitreous humor of the eye, and are vital to the normal development of bones and teeth. These fibrils are formed from highly ordered 67 nm staggered arrays of collagen molecules, producing the characteristic D-periodic structure of fibrillar collagens (S. Fig. 1). Because the length of the type II collagen molecule (~300 nm) is not an integer multiple of 67 nm, the fifth molecular segment does not extend across the whole D-period. Therefore a gap region is defined, where there are only four molecular segments and an overlap region where there are five. This arrangement is known as the Hodge-Petruska scheme (see packing / microfibril in S. Fig. 1) (1). Each D-period is composed of arrays of five fragments of five different collagen molecules, which between them contain the complete sequence of one collagen molecule. The type II collagen molecule has three identical \(\alpha_1\) polypeptide chains of 1060 amino acid residues each, with a large uninterrupted triple-helical region and relatively short, non-helical telopeptides (19 amino acid residues in the N-telopeptide, and 27 amino acid residues in the C-telopeptide) (2) that do not possess the Gly-Xaa-Yaa repeating primary structure found in the triple-helical region. The length of the \(\alpha\)-chains is the same, but they are displaced from one another by one residue in the triple-helix, to allow for its proper super-coiling (2).

\textit{In vitro}, the polymerization of collagen molecules into fibrils is an entropy driven self-assembly process (3), whereas in living tissues it
also involves cellular and specific extracellular matrix (ECM) interactions (4). Proteoglycans (PG's), such as decorin, fibromodulin, and biglycan, bind type II collagen fibrils to stabilize the larger fibril-bundles (fibers composed of multiple fibrils). The diameter of the latter being regulated through interactions with the PG's protein core and anionic glycosaminoglycan (AGAG) chains (5). However, the fibrils internal order and stability are chiefly maintained by intermolecular crosslinking between collagen molecules. Intra-fibrillar links are formed through covalent bonds between lysine and hydroxylysine (Hyl, a modified amino acid specific to collagen, single letter abbreviation U) through a Schiff base reaction (6, 7). These bonds are formed between Lys and Hyl residues within both the helical and non-helical telopeptide domains of collagen molecules within each D-period. Thus, it is apparent that although the telopeptides do not have triple-helical structure, they are very important for collagen crosslinking and the assembly of collagen molecules into fibrils (8). Proper crosslink formation ensures the structural integrity and stability of fibrils and the tissue. Abnormal crosslinking, either enhanced or inhibited, leads to connective tissue diseases such as Ehlers-Danlos syndrome, Marfan syndrome, glaucoma and skin fragility in normal aging (9). Furthermore, the proper conformation of the telopeptides has been shown to be important for normal fibril structure (10-13). Accurately defining the structure of the D-periodic, Hodge-Petruska (1) arrangement of type II collagen, that is: gap and overlap ratio, telopeptide conformations and crosslink locations, may facilitate our understanding of these conditions.

Structural studies of type II collagen packing has been complicated by the technical limitations its fibrillar nature presents. The insolubility and size of mature collagen molecules makes nuclear magnetic resonance (NMR) invisible and only small fragments of the collagen molecule or synthetic, short collagen-like peptides have been crystallized and studied by (single-crystal) X-ray crystallography so far (14). Furthermore, disruption of the fibril structure in order to study the constituent collagen molecules in isolation compromises the goal of understanding the molecular packing within fibrils and the native collagen structure within the D-period. Electron microscopy offers significant capabilities for studying fibrillar collagens, but does not provide sufficient resolution for characterizing the telopeptide conformation, and sample preparation may disturb subtle, small features such as the telopeptides.

Type II collagen fibrils have an intrinsically crystalline D-periodic structure, even though their lateral arrangement is less well ordered (15, 16). This means that X-ray diffraction experiments on whole intact tissue samples can provide information concerning the axial organization of collagen. Such structural data have been obtained previously (15, 16), but were less detailed than that seen recently for collagen type I (obtained by our group) (17-19). In this study, we applied techniques developed during earlier investigations of type I collagen structure (17, 19-21), to better define the D-periodic structure of type II collagen fibrils in situ. We present here the axial D-periodic structure of native type II collagen in lamprey notochord, determined by Multiple Isomorphous Replacement (MIR) to 1.9 nm resolution.

**EXPERIMENTAL PROCEDURES**

**Materials-** Adult female lampreys were kindly provided by the Ludington Biological Station of the U.S. Fish and Wildlife Service and J. Ellen Marsden of the University of Vermont.

**Fiber diffraction-** The notochords of adult lampreys were carefully harvested and dissected from their sheaths, nearby muscle and other tissues, and stored at pH 7.5 in Phosphate Buffered Saline (PBS, 8 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, 137mM NaCl) or Tris Buffered Saline (TBS, 25 mM Tris, 150 mM NaCl) for 1 to 6 h at 4°C before each experiment. The notochord samples were cut longitudinally into pieces (1 x 2 x 25 mm) along their longitudinal axis and mounted in custom made sample holders that provided for 1-2% stretching and preservation of their hydrated state. Isomorphous staining was performed as previously described (17, 20). Briefly, samples were treated with iodine (KI) and platinum (PtCl$_4$) for 1 h, followed by 15 min of washing in a large excess of buffer. I$^-$ ions covalently bind to tyrosine and in some cases histidine, whereas Pt$^{3+}$ ions associate with the sulfur group of methionine.
X-ray diffraction studies were performed primarily using the small and wide angle fiber diffraction instruments on the BioCAT Beamline 18ID at the Advanced Photon Source (APS), Argonne National Laboratory, IL, USA. Some preliminary studies used the wide angle instrument on the BioCARS beamline 14BMC at the APS. The sample-to-detector distance was 2000 mm for SAXS (1-7 meridional orders) and 150 mm for WAXS and microfocus WAXS (5-37 meridional orders, medium to high resolution fiber crystallography), and the wavelength was 1.033 Å. Diffraction images were recorded on a custom made ‘Brandeis’ detector (48 x 48 micron pixel size) (22-24)) for WAXS, and on an AVIEX PCCD 16080 detector (38.4 x 38.4 micron pixel size) (22) for SAXS. The exposure time was limited to 1 second to safe-guard against radiation/heat damage to the portion of the sample being shot.

Data analysis- Diffraction images were processed using CCP13 Fibrefix (25) and Fit2D (26) software for background subtraction, and the CCP13 software for measurements of meridional intensities (17). Measured intensities (S. Fig. 2), (squares of amplitudes) were used for calculation of difference Patterson maps and a one-dimensional electron density map of the collagen type II axial unit cell (the D-period) by reverse Fourier transformation (in-house software). The scaling of the experimental amplitudes and the solution of the phases were performed as described previously (17, 20, 19).

Lamprey collagen type II amino acid sequence- The entire collagen type II sequence is known for human, rat and mouse and partial sequences are known for lamprey (the primary source for this study) [ExPASy sequence data bank codes P02458, P05539, P28481, Q218Y0 and Q218X9]. Unlike collagen type I, collagen type II is highly homologous across the species (27), and its full sequence may therefore be reasonably estimated by homologous comparison (N-terminal half from rat, mouse and human; C-terminal half from the lamprey sequence, Q218X9) (28).

Model electron density map- The model electron density map was calculated (17, 20), using the residue level scattering factors of Hulmes et al. (21), the amino acid sequence described previously, the Hodge-Petruska scheme, and assuming a regular unit height spacing of the molecular helix (residue to residue distance). The telopeptides were initially modeled as elongated structures extending straight from the ends of the triple-helical region. After comparison with the difference Patterson and Fourier maps, the telopeptides were modeled as folded structures, similar to that of type I collagen C-telopeptide (although the N-telopeptide is required a double folded conformation to fit the experimental data, see Results).

Determination of the gap/overlap ratio- In this study, we took the approach that the gap/overlap ratio is determined by the extents of the telopeptides (17) and along the lines of the height/depth ratio approach of Bradshaw et al. (20) Hence the overlap start and end points were determined by considering:

1) The position of the iodine peaks for the telopeptides that give the location of the telopeptide ends. The distance from the telopeptide termini to the fold can be estimated from the number of known amino acids spanning the distance.

2) The point of the overlap region above average height versus gap region below average depth. The low resolution of the study means that the influence of the first few orders (particularly the first) produces a tail effect that makes the overlap appear larger than it is. This was partly countered through the calculation of the native electron density map with the amplitude of the first order reduced ten times (S. Fig. 3).

3) The model parameters that fit all of the following: difference Fourier heavy atom positions, difference Patterson peak separation values, and best fits the native electron density, provided there is compliance with points 1 and 2.

4) Examination of the TEM images of type II fibrils to confirm that the values determined from 1-3 were reasonable. We hold that the X-ray data are the most accurate approach due to the use of native samples, better resolution and due to the multiple lines of approach used to reach it.

Transmission electron microscopy (TEM)- Collagen type II fibrils were studied at the University of Chicago Electron Microscopy Center. Images were examined under 300 KV using an FEI Tecnai F30 microscope with a Gatan CCD digital micrograph (4kx4k) as detector. Sample preparation was the same as for the diffraction experiments, followed by fixation,
embedding and sectioning into 90 nm sections on a Reichert-Jung Ultracut E microtome. Sections were stained with uranyl acetate and lead citrate to enhance the image contrast. We found that suspensions greatly enhanced the clarity of the human type II collagen fibrils. For these preparations, cartilaginous tissues were homogenized in TBS on ice for 1 min to release collagen fibrils and other matrix components. These samples were then placed on grids and stained with uranyl acetate.

The D-period from the X-ray structure was confirmed against the TEM data by sampling the gap/overlap ratio in the clearest TEM image (i.e. such as that shown in S. Fig. 4F) over ten unit cells. The standard deviation was 0.0049 and the percentage error (as measured by the mean discrepancy between the sum of gap and overlap measurements versus the single D-band measurement was <0.05%), arriving at a value of 0.419 : 0.579 D for the gap:overlap. The scale bars showing the gap-overlap and D-band extents were then superimposed on the other TEM images (see Results).

Telopeptide structure prediction and minimization- The extent of folding or compression required of the telopeptides to fit the native electron density map was determined from the difference Patterson and Fourier maps. In addition, structure prediction calculations were performed for the conformation of the telopeptides. Both N- and C-termini were examined using the Chou-Fasman (29) and SOPMA (30) structure prediction methods, and the results were compared with the diffraction and electron microscopy data (D-period gap/overlap ratio, position of heavy atom peaks). The results were found to be in good agreement; therefore atomic coordinates of the telopeptides for one alpha chain were generated de novo from the amino acid sequence and energy minimized using NAMD (31) as a further check of the hypothesis that the telopeptides are folded.

Molecular model and surface rendering- The molecular surfaces were calculated using ‘spock’(32-34) with the default options, except the surface polygon parameter which was set to 120 for improved surface definition. The display option was set to 'mesh' to allow the underlying bonds of the model to be seen. The calculated mesh is not an electron density map, simply a rendering of the molecular surface.

RESULTS

X-ray diffraction data. A medium-wide angle diffraction pattern from native (unstained) type II collagen is shown in Fig. 1, with the central section containing the meridional series indicated. Diffraction images from native lamprey notochord and two heavy-atom derivatives were recorded, showing meridional reflections up to the 37th order or ~18 Å (integrated and scaled intensities of orders 1-35 for native samples and derivatives are presented in S. Fig. 2).

Location of heavy-atom-binding sites and difference Fourier and Patterson maps. The amplitudes of the native samples were subtracted from the amplitudes of the derivative samples and the resulting differences for each order, together with the calculated phases were used in a reverse Fourier transformation. This operation produces an electron density map, whose peaks correspond to the positions of heavy atoms within the native-like unit cell (the D-period) of the collagen type II fibril. The iodine derivative difference Fourier map is shown in Fig. 2. Although the majority of these heavy atom binding positions appear to correspond to the expected collagen amino acid residue labeling positions, some appear to have bound isomorphously at an unexpected location (no liable amino acid residues from the collagen sequence), and given the corresponding native electron density, this may indicate the presence of the small leucine rich repeat protein (sLRRP), biglycan, or a related sLRRP at this location. It is known that sLRRP's are attached to the surface of type II collagen fibrils at around 0.7-0.8D (see discussion).

The presence of tyrosine residues in the telopeptides and telopeptide region, allows the detection of the ends of the telopeptides (difference Fourier, Fig. 2B, and Patterson maps Fig. 2C and 3D). From these data it was clear that the telopeptides are highly contracted or folded (Figs 3 and 4), a possibility supported by the native electron density map (see below and Fig. 2A).

Collagen type II one-dimensional electron density profile. The native electron density map
was calculated using reverse Fourier transformation of the experimentally derived structure factors (Fig. 2). This map represents the sum of the five collagen molecular segments in lateral projection (perpendicular to the fiber axis) in the Hodge-Petruska scheme (S. Fig. 1), and is a direct function of the axial D-periodic structure of the type II collagen fibril. This map allows deductions to be made concerning the conformation of the collagen molecules, the collagen fibril parameters, and the telopeptide conformations (in conjunction with the heavy atom labeling data).

From the one-dimensional electron density map, the overlap is measured to be 0.42D, which is shorter than that of the collagen type I fibril, and the gap is approximately 0.58D (Fig. 2). This overlap to gap ratio of 0.42:0.58 is confirmed by the TEM data (Fig 5 and S. Fig. 4). Compared with type I collagen, these parameters might appear confusing, because the type II molecule is four amino acid residues longer than the type I molecule, which would suggest that the overlap to gap ratio should be closer to 1:1 than the type I ratio of 0.46:0.54 (17, 35). After considering the difference Patterson and Fourier data, it is reasonable to assume that the shorter collagen type II overlap region is a consequence of the telopeptides adopting a conformation that makes them shorter in linear projection (see S. Fig. 3), (17). The telopeptides would have to be either highly compressed (‘contracted’) or tightly-turned to fit the measured gap-overlap ratio and difference Fourier and Patterson data. The compressed model does not seem plausible due to peptide bond constraints; compressed telopeptides would require a height translation of ~0.8 Å to fit within the 0.42D overlap region.

One-dimensional model of collagen type II D-periodic structure and structure prediction. To aid the analysis and interpretation of the one-dimensional electron density map, an electron density model of one D-period was calculated from the collagen type II amino acid sequence and the individual whole amino acid residue structure factors (Fig. 2) (21). In the model, the collagen molecules are staggered by 234 residues (~67 nm) and the individual peptide α-chains are staggered by one amino acid residue within each triple-helix segment (17, 19). The initial model assumed a conformation of the telopeptides where they were straight and relaxed (Figs. 3 and 4, and S. Fig. 3). This model provided parameters for the D-period that disagreed with the one-dimensional map obtained from the experimental data (S. Fig. 3 and Fig. 2). The model overlap region (0.53D) was longer than the experimentally determined value (0.42D) and the gap region was shorter (theoretical 0.47D, experimental 0.58D).

Assuming that the telopeptides may form turns, unstructured coils and relaxed strands (17, 21, 36), their amino acid sequences were examined and possible conformations were predicted using the Chou-Fasman and SOPMA methods (S. Tables 1, 2). Both these methods and the difference Fourier and Patterson data suggested tight-turns in the N- and C-telopeptides, which would change the gap/overlap ratio in the model to better fit the experimentally determined values.

Telopeptide conformations. According to the structure predictions (S. Table 1, Fig. 3), the N-telopeptide can form two turns, one at position 11G-12G and another at position 4G-5G. The telopeptide section between the triple-helix and the first turn has greatest propensity to form β-strands. The peptide chains of the triple-helix itself are considered to be in almost a relaxed β-strand conformation, which is close to the polyproline II helical conformation. Given that the telopeptide is directly connected to the triple-helical regions, the N-telopeptide may have some ‘structural memory’ (17, 37) of the polyproline II helix, despite not having a triple-helical conformation until it makes a 180° turn. A second turn is predicted after a straight section. The fragment between the two turns and the rest of the telopeptide after the second turn has a greater propensity to form a ‘random coil’ (or turn), forming a high electron density area relative to the rest of the triple-helix, between amino acid positions 11 and 15 (Fig. 3).

This N-telopeptide conformation makes the overlap region shorter and also brings the Hyl9 residue closer for crosslinking to the Lys949 in monomer five (a close molecular packing neighbor in type I collagen). If the N-telopeptide was simply straight and relaxed, Hyl9 would be too distant from Lys949 and the alternative crosslinking partners Lys238 (monomer 2), Lys703 (monomer 4) and Lys937 (monomer 5) are even further distant. The average spacing between collagen type II monomers in the fibril is about 1.3 nm, based on equatorial diffraction data (not
shown) and collagen type I parameters (19). There are eight C–C and two C–N bonds in the lysyl-hydroxylysine crosslink between the two peptide backbones. The C–C bond is 1.54 Å and C–N bond is 1.47 Å (38). Hence, the length of the whole lysyl-hydroxylysine link from peptide backbone to peptide backbone maybe around 1.5 nm at a maximum. Therefore both the lysine and the hydroxylysine have to be in approximately the same lateral plane, or no more than 0.8 nm apart in the axial direction (within three amino acid residue positions) in order to form the crosslink.

The C-telopeptide shows a propensity to form one sharp turn at the Gly13 position of the telopeptide (Gly1046 of the full molecular sequence) (S. Table 2 and Fig. 4). The fragment of the telopeptide between the turn and triple-helical region has a predicted structure close to a relaxed helix due to methionine clusters (S. Table 2), disturbed by glycine residues. The rest of the C-terminus after the turn is more likely to be a ‘random coil’ (S. Table 2, Fig. 4). This folded conformation causes additional shortening of the overlap and extension of the gap region in the electron density model. It also moves Lys1050 by approximately eight amino acid positions, moving it much closer to Hyl106 for covalent crosslinking (Fig. 4) (see previous discussion of Lys-Hyl crosslink length).

The structures of the collagen type II telopeptides, as determined from the experimental data (difference Patterson and Fourier maps) and supported by the predicted conformations, were incorporated into the refined one-dimensional model of the D-periodic molecular packing (Fig. 2 and S. Fig. 3a). This model electron density map provided a much better approximation to the experimental data. Structure factors were calculated from the refined model and used to generate a model electron density map at the same resolution as the experimental map (35 meridional orders, Fig. 2). This allowed the simple R-factor (19) to be calculated, as an estimate of error of the goodness of fit between the two maps. The calculated R-factor was 0.19.

TEM data: mammalian and lamprey fibrils are homologues but tissue architecture differs. Micrographs of collagen type I (rat-tail tendon, human quadriceps tendon) and type II fibers (lamprey notochord, human articular cartilage, bovine articular cartilage) were examined (Fig. 5 and S. Fig. 4). The collagen fibers in these images show the typical pattern for fixed collagenous tissues: black and white bands with approximately 64 nm periodicity (the D-period is shortened during fixation and embedding (15, 39). These images also confirm previous observations that collagen type II fibrils from lamprey notochord do not show any detectable differences from those of mammalian tissues (15, 16). The lamprey notochord fibrils have similar diameters of about ~35 nm and the typical positive staining pattern observed for mammalian fibrils (Fig. 5 and S. Fig. 4).

Despite the similarities between fibril morphology, the architecture of collagen fibril arrangement differs between mammalian and lamprey tissues. There are also some differences in cellular and PG content in these tissues. Lamprey notochord has a very specific cell distribution and most cells were mechanically removed during sample preparation, whereas cartilage cells are embedded in the collagen meshwork and can be seen throughout the whole tissue (40, 41). Lamprey notochord principally contains the PG's, biglycan types I and II, which are similar in sequence to the bovine and human biglycans (comparison of lamprey sequence fragments Q9DE00 and Q9DDZ9 with human P21810 and bovine P21809, (42)), and hence are most likely to be structurally related to decorin and fibromodulin. In contrast, mammalian cartilage contains several different types of PG's, glycoproteins and other types of collagen present (27, 42). The specific composition of the ECM plays a crucial role in the formation of collagen fibrils and the organization of the fibrillar meshwork. This influence appears to give rise to the complexity of the articular cartilage matrix, as observed by TEM (S. Fig. 4d), in comparison with the more ‘primitive’ architecture of notochord tissue (S. Fig. 4b).

DISCUSSION

TEM images of collagen type I and type II fibers from the same tissues used in X-ray diffraction experiments, provided complementary information that correlated well with the results of the X-ray diffraction experiments and with the derived one-dimensional model of collagen type II’s D-period structure (S. Fig 3 and Figs 2, 5).
The experimental one-dimensional electron density map (Fig. 2) indicated that a previous observation reported for type I collagen (19, 20, 43) also applies to type II collagen: there are 234 amino acid residues in the molecular segments within the Hodge-Petruska D-periodic fibril packing structure, which is 67 nm long in hydrated samples, and 64-65 nm long in dehydrated, fixed samples (Fig. 5). However, the overlap is 0.42D +/- 0.03 (28.14 nm or 98 amino acid residues long) and the gap is 0.58D (38.86 nm or 136 amino acid residues) in type II collagen, whilst in type I collagen, the overlap is 0.46D and gap is 0.54D. The overlap/gap ratio for type II collagen has previously been suggested to be closer to 0.4D/0.6D than type I (39, 40, 44, 45) and is also confirmed from the TEM data (Fig. 5 and S. Fig. 4). This indicates that although there are some obvious similarities, collagen types I and II differ significantly at the level of molecular packing. These differences may arise due to the specific telopeptide conformation of each collagen type.

**Experimental and model electron density maps of the type II collagen D-period.** Fig. 2 shows both the one-dimensional map of the collagen type II D-periodic structure (blue) and the model constructed from the amino acid sequence and residue scattering factors (green). These maps have similar features, although some peaks are not common to both. The initial theoretical model and experimental peak density positions differ significantly, indicating that the real molecular conformation is not simply linear, particularly in the configuration of the telopeptides (Figs. 2-4) and S. Fig 3). A major peak in the gap region of the electron density map for native type II collagen that is not present in the model (see below), may correspond to dermatan sulfate PG (DSPG) binding sites, which in the case of lamprey notochord is biglycan or biglycan-like protein, see above (40) This is in agreement with the principal attachment sites for decorin in the e and d band of type I collagen, as determined by TEM (5, 35, 47). Other peaks on the map may correspond to electron dense side-chain residues of the collagen molecules or other ECM molecules ligated to the surface of the collagen type II fibril, and/or differences in the amino acid residue-to-residue spacing. The latter may in part arise from differences in helical symmetry (e.g. 10/3 vs. 7/2 symmetry). Other significant differences between the experimental and model maps may be explained by the presence of other PG's, fibronectin and aggrecan molecules that may be associated with collagen fibers in lamprey notochord. Some of these peaks may be present in the type I collagen electron density map (17), but if present, are much more subtle. Their greater prominence in the type II collagen one-dimensional map is due to the fact that type II fibrils are much thinner then type I fibrils (~35 nm and 100-200 nm, respectively, although larger fibers, i.e. bundles of fibrils, may also be present), thereby increasing the effective occupancy and electron density contrast of PG molecules attached to the outside of the fibril.

**Informative differences between model and experimental electron density maps.** The model of the collagen type II D-periodic structure has several peaks in both the gap and overlap regions that correspond to clusters of amino acids with electron dense side chains, and in general these peaks are also present in the experimental electron density map. However, there are two major differences between the experimental electron density map and the model: a trough in the map at 0.275D and a peak at 0.8D that are not readily explained by the collagen amino acid sequence or electron scattering density alone. As already noted, the data for the experimental electron density map were obtained from native tissues, which contain not only collagen, but also other ECM molecules. Biglycan may bind to several sites on the collagen monomer, but the electron density map indicates that one of these attachment sites on fibrillar type II collagen at 0.8D (905G – 915P, monomer four) is occupied in a highly ordered fashion. This is in agreement with the principal attachment sites for the biglycan homologue, decorin, in the e and d staining bands of type I collagen (48, 49). The presence of the model peak, in contrast to the experimental electron density map trough at 0.275D may suggest a significant and common molecular inflection amongst the five molecules in the overlap region, before and after this position. Such that the molecular sections passing through 0.275D are straight but tilted on either side of this position, accounting for the greater electron density on either side of this point and the lower relative electron density at 0.275D itself. This type of inflection is also present in the type I collagen microfibril (19). Neither of these two differences
between the experimental map or model affect the measurement of the gap-overlap ratio however.

Telopeptide conformations. The conformations of the telopeptides are important for collagen fibril assembly due to the covalent Lys-Hyl crosslinks formed between the collagen molecules within these non-helical domains (50, 51). Lysine or hydroxylysine is deaminated by lysyl oxidase and a covalent link is formed spontaneously; this may occur at two places in the collagen type II molecule (7) (Figs. 3 and 4). The distribution of Lys and Hyl residues within the collagen molecule (Figs. 3 and 4) determines the pattern of crosslinks and therefore the molecular stagger within the fibril. The N- and C-telopeptides are involved in crosslinking and their conformations affect the axial position of the crosslinking residues, Hyl9 and Lys1050, which bind to Lys949 and Hyl106 in the triple-helical region respectively. Thereby influencing crosslink formation and therefore fibrillogenesis.

Previous investigations of the structure of fibrillar collagen indicated that the telopeptides form contracted and folded structures (9, 17, 21, 39). Electron microscopy studies of human collagen type II fibers suggested that the folding of the telopeptides is similar to that proposed here (39), but the 20 nm resolution of those studies was not sufficient to provide conclusive evidence (the stretched N- and C-telopeptides are 4.5 nm and 5.4 nm in length, respectively). Similarly, studies with synthetic or ex vivo collagen N- and C-telopeptides indicated an axial compression of the fragment/s, with possible folding (52-54). However, the specific telopeptide conformations were not determined. Subsequently, in X-ray diffraction investigations of native rat-tail tendon collagen type I in situ, we showed the one-dimensional packing structure of the collagen molecules in the D-period. The conformation of the C-telopeptides showed a sharp turn around residues Pro13 and Gln14 and the specific locations of the lysine-hydroxylysine crosslinks (17). These results were later confirmed when the full three-dimensional, higher resolution structure was determined (19). In this study the conformations of collagen type II telopeptides were inferred from the native and difference Fourier density maps, and predicted by the Chou-Fasman method and confirmed by SOPMA (S. Tables 1 and 2). Both telopeptides have a propensity to form a sharp turn in the middle of their sequences (Figs. 3 and 4), which make the whole collagen molecule 2% (5.7 nm) shorter and changes the gap/overlap ratio of the model. These predictions agreed well with the observation that the overlap region must be shorter than linear telopeptide structures would allow, and also with the peak distributions in the difference Patterson and Fourier maps, since the heavy atom binding locations included positions within the telopeptides themselves.

Significantly, the N-telopeptide structure of type II collagen appears to differ from that of type I collagen, which does not have any sharp turns (17, 19). In contrast, the proposed collagen type II C-telopeptide structure is much closer to that observed for the collagen type I C-telopeptide and to the prediction of Ortolani (39). It has tyrosine (Tyr1058) and methionine (Met1055) residues at the end that have been moved deeper into the overlap region, closer to the N-terminus. This conformation is also supported by the difference Patterson and Fourier maps.

Implications of telopeptide structure for crosslinking. The derived conformation of the collagen type II telopeptides suggests a specific, well ordered crosslinking pattern in the fibril that is different from that of collagen type I. It may also be the molecular basis of the differences in fibril diameter and fibril bundle organization between two of the most prominent fibrillar collagens, types I and II. The type II collagen molecule has more candidate residues for crosslinking than the type I molecule. At the C-terminus, there are three potential crosslink forming Lys residues, due to the homotypic chain composition of type II molecules. In contrast, crosslinking at the C-terminus of the type I molecule appears to occur primarily through its two alpha 1 chains. Similarly, at the N-terminus of collagen type II, Hyl9 may form three covalent bonds with Lys residues on monomer five, whereas the potential for N-terminus crosslinking in collagen type I is diminished in comparison (19). Therefore, the greater potential for crosslinking in collagen type II may lead to greater numbers of supporting covalent bonds within and possibly even between the fibrils (within fibril bundles) than in collagen type I. Thus, the type II fibril is likely to be more stable than the collagen type I fibril due, in part, to a higher crosslink
content. The greater potential for interfibrillar crosslinking in type II collagen may also result in a more complex and possibly more stable network-like tissue organization. At the same time, because the type II N-telopeptide is significantly more bulky then type I collagen’s, it may also function in sterically inhibiting the formation of type II fibrils as large as those observed for type I. This in turn would lead to more loosely organized fibril bundles (or thick fibrils) composed of these small and relatively thin fibrils (55).

Acknowledgments - We thank the Ludington Biological Station of the U.S. Fish and Wildlife Service and J. Ellen Marsden of the University of Vermont for the donation of lamprey specimens, Thomas Schmid and Vincent Wang of Rush University Department of Sports Medicine for the donation of cartilage samples. We also thank Yimei Chen of the Electron Microscopy Center at University of Chicago for help in sample preparation and collection of TEM data. Special thanks to Raul Barrea for his hard work in the development of microfocus capabilities at BioCAT and to the staff of both the BioCAT and BioCARS groups. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-ENG-38. BioCAT is a National Institutes of Health-supported Research Center RR-08630. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR-007707. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Center for Research Resources or the National Institutes of Health. This work was also supported by the National Science Foundation (Grant #MCB-0644015 CAREER).
REFERENCES

1. Petruska, J. and Hodge, A. (1964) Proc Natl Acad Sci U S A 51, 871-6
2. Prockop, D. J. and Kivirikko, K. I. (1995) Annu. Rev. Biochem 64, 403-434
3. Kadler, K., Hojima, Y., and Prockop, D. (1987) Journal of Biological Chemistry 262, 15701, 15696
4. Kadler, K. E., Hill, A., and Canty-Laird, E. G. (2008) Curr. Opin. Cell Biol 20, 495-501
5. Scott, J. E. and Haigh, M. (1988) Biochem. J 253, 607-610
6. Prockop, D., Berg, R. A., Kivirikko, K., and Uitto, J. (1976) “Intracellular steps in the biosynthesis of collagen” in Biochemistry of Collagen., Plenum Press, New York
7. Robins, S. P. and Duncan, A. (1983) Biochem. J 215, 175-182
8. Helseth, D. L. and Veis, A. (1981) J. Biol. Chem 256, 7118-7128
9. Eyre, D. R., Paz, M. A., and Gallop, P. M. (1984) Annu. Rev. Biochem 53, 717-748
10. Bank, R. A., Robins, S. P., Wijmenga, C., Breslau-Siderius, L. J., Bardoeil, A. F. J., Sluijs, H. A. V. D., Pruis, H. E. H., and TeKoppele, J. M. (1999) Proc Natl Acad Sci U S A. 96, 1054–1058
11. Gineyts, E., Cloos, P. A., Borel, O., Grimaud, L., Delmas, P. D., and Garnero, P. (2000) Biochem. J 345 Pt 3, 481-485
12. Pieper, J. S., van der Kraan, P. M., Hafmans, T., Kamp, J., Buma, P., van Susante, J. L. C., van den Berg, W. B., Veerkamp, J. H., and van Kuppevelt, T. H. (2002) Biomaterials 23, 3183-3192
13. Garnero, P., Schott, A., Prockop, D., and Chevrel, G. (2009) Bone 44, 461-466
14. Brodsky, B. and Persikov, A. V. (2005) Adv. Protein Chem 70, 301-339
15. Eikenberry, E., Childs, B., Sheren, S., Parry, D., Craig, A., and Brodsky, B. (1984) J Mol Biol 176, 261-77
16. Brodsky, B., BelBruno, K., Hardt, T., and Eikenberry, E. (1994) J Mol Biol 243, 38-47
17. Orgel, J., Wess, T., and Miller, A. (2000) Structure Fold Des 8, 137-42
18. Orgel, J., Miller, A., Irving, T., Fischetti, R., Hammersley, A., and Wess, T. (2001) Structure (Camb) 9, 1061-9
19. Orgel, J., Irving, T., Miller, A., and Wess, T. (2006) Proc Natl Acad Sci U S A 103, 9001-5
20. Bradshaw, J., Miller, A., and Wess, T. (1989) J Mol Biol 205, 685-94
21. Hulmes, D., Miller, A., White, S., and Doyle, B. (1977) J Mol Biol 110, 643-66
22. Fischetti, R., Stepanov, S., Rosenbaum, G., Barrea, R., Black, E., Gore, D., Heurich, R., Kondrashkina, E., Kropf, A. J., Wang, S., Zhang, K., Irving, T. C., and Bunker, G. B. (2004) J Synchrotron Radiat 11, 399-405
23. Barrea, R., Davidson, M., Orgel, J., Weng, T., Kondrashkina1, E., Gore, D., Heurich, R., Vukonich, R., and Irving, T. (2005) “The 18ID BioCAT microprobe for X-ray imaging, microXAFS and microdiffraction studies on biological samples” in Spring-8, 2005 users meeting.
24. Phillips, W. C., Stewart, A., Stanton, M., Naday, I., and Ingersoll, C. (2002) J Synchrotron Radiat 9, 36-43
25. Rajkumar, G., Al-Khayat, H. A., Eakins, F., Knupp, C., and Squire, J. M. (2007) J Appl Crystallogr 40, 178-184
26. Wess, T., Hammersley, A., Wess, L., and Miller, A. (1995) J Mol Biol 248, 487-93
27. Kadler, K. (1995) Protein Profile 2, 491-619
28. Zhang, G., Miyamoto, M. M., and Cohn, M. J. (2006) Proc Natl Acad Sci U S A 103, 3180-3185
29. Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222-245
30. Geourjon, C. and Deléage, G. (1995) Comput. Appl. Biosci 11, 681-684
31. Phillips, J., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R., Kale, L., and Schulten, K. (2005) J Comput Chem 26, 1781-802
32. Christopher, J., Swanson, R., and Baldwin, T. (1996) Comput Chem 20, 339-45
33. Perumal, S., Antipova, O., and Orgel, J. (2008) Proc Natl Acad Sci U S A 105, 2824-9
34. Orgel, J. P. O., Eid, A., Antipova, O., Bella, J., and Scott, J. E. (2009) PLoS ONE 4, e7028
35. Schmitt, F. O., Cecil E. Hall, and Marie A. Jakus (1942) Journal of Cellular and Comparative Physiology 20, 11-33
36. Hulmes, D., Wess, T., Prockop, D., and Fratzl, P. (1995) *Biophys J* **68**, 1661-70
37. Yu, E. W. and Koshland, D. E. (2001) *Proc. Natl. Acad. Sci. U.S.A* **98**, 9517-9520
38. *Weast, RC* (ed.): *Handbook of Chemistry and Physics*, 66th ed., CRC Pres, Boca Raton, FL
39. Ortolani, F., M. Giordano, and M. Marchini (2000) *Biopolymers* **54**, 448-463
40. Pasteels, J. (1958) *Développement des agnathes*. In: Grassé P, editor. *Traité de Zoologie*, Masson et Cie, Paris
41. Mow, V. and Ratcliffe, A. (1997) *Structure and function of articular cartilage and meniscus*, Raven Press, New York
42. Shintani, S., Sato, A., Toyosawa, S., O'hUigin, C., and Klein, J. (2000) *J. Mol. Evol* **51**, 363-373
43. Orgel, J. P., Wess, T. J., and Miller, A. (2000) *Structure* **8**, 137-42
44. Bos, K., Holmes, D., Kadler, K., McLeod, D., Morris, N., and Bishop, P. (2001) *Journal of molecular biology* **306**, 1011-1022
45. Ortolani, F. and Marchini, M. (1993) *Boll. Soc. Ital. Biol. Sper* **69**, 107-113
46. Park, H., Huxley-Jones, J., Boot-Handford, R. P., Bishop, P. N., Attwood, T. K., and Bella, J. (2008) *BMC Genomics* **9**, 599
47. Scott, J. (1993) “Proteoglycan-fibrillar collagen interactions in tissues: dermatan sulphate proteoglycan as a tissue organiser” in *Dermatan Sulphate Proteoglycans*, Portland Press, London, 165-183
48. Scott, P., Winterbottom, N., Dodd, C., Edwards, E., and Pearson, C. (1986) *Biochemical and biophysical research communications* **138**, 1348-54
49. Scott, J. and Glanville, R. (1993) *Biochemical Society transactions* **21**, 123S
50. Martin, G., Gross, J., Piez, K., and Lewis, M. (1961) *Biochim Biophys Acta* **53**, 599-601
51. Piez, K., Lewis, M., Martin, G., and Gross, J. (1961) *Biochim Biophys Acta* **53**, 596-8
52. Otter, A., Scott, P. G., and Kotovych, G. (1988) *Biochemistry* **27**, 3560-3567
53. Otter, A., Scott, P. G., and Kotovych, G. (1993) *Biopolymers* **33**, 1443-1459
54. Liu, X., Otter, A., Scott, P. G., Cann, J. R., and Kotovych, G. (1993) *J. Biomol. Struct. Dyn* **11**, 541-555
55. Holmes, D. and Kadler, K. (2006) *Proc Natl Acad Sci USA* **103**, 17249-17254
FIGURE LEGENDS

**Fig. 1.** Medium-wide angle X-ray diffraction pattern of collagen type II fibrils from lamprey notochord with 15-20 Å resolution of meridional reflections series.

**Fig. 2.** Experimental and model electron density maps of native type II collagen D-periodic axial structure (A). Difference Fourier map (B): positive peaks show the attachment positions of iodine heavy atoms within the D-period, 0% is towards the beginning (N-terminal end, beginning of overlap) of the D-period, and 100% is towards the end (end of gap region). The positions of the iodine labile residues are shown in black font for the refined model. Red font shows the positions of tyrosine residues were the telopeptides to adopt an extended, linear conformation. Difference Patterson function for iodine derivative (black), and model Patterson function (red) based on the iodine labeling positions determined from A (C). Difference Patterson function as for C, and model Patterson function with the telopeptides adopting an extended and linear conformation (D). Note: The scaled amplitudes (S. Fig. 2) were also used to calculate difference Patterson maps for the iodine derivative). The peak position in a difference Patterson plot corresponds to the distance between two points within the unit cell (plotted from 0 to 0.5 unit cell length in fractional coordinates). The peak height is a function of the relative electron density at the termini of each distance. In the present case, a peak represents the distance between heavy atoms that have labeled the unit cell contents isomorphously.

**Fig. 3.** Modeling of the N-telopeptide of collagen type II: in its extended conformation (A), possible cross-linking pattern of hexagonally packed type II collagen monomers with the N-telopeptide carrying monomer 1 marked yellow (B), folded telopeptide conformation (C) supported by hydrogen bonds; Lys residue is shown in blue. The Hyl shifts towards the triple-helix (to the right) making the possibility of forming a covalent bond with Hyl947 of monomer 5 more favorable (D, two α-peptides are shown extended, blue, one is shown folded, green); the crosslink is indicated by light-red arrow. Black bars alongside blue bars in d show the large axial distance crosslinks would need to reach if the telopeptide is extended (blue, see also A). Red bar shows short axial distance spanned by folded telopeptide (green, see also C). Monomers are approximately 1.3 nm distant from each other in lateral space (B). Hydroxylysine is defined as single letter code “U”. Mesh displayed around model shown parts A and B is a molecular surface rendering to show the model outline (see methods), not electron density.

**Fig. 4.** Modeling of the C-telopeptide of collagen type II: in extended (A) and folded (B) conformations, supported by hydrogen bonds; Hyl residue is shown in blue. The possible cross-linking pattern of hexagonally packed type II collagen monomers is shown with the C-telopeptide carrying monomer 5 marked in yellow (C). The Hyl shifts towards the triple-helix (to the left) making its covalent bonding with Lys106 of monomer 1 more favorable (D, two α-peptides are shown extended, blue, one is shown folded, green); the crosslink is indicated by light-red arrow. Black bars alongside blue bars in d show the large axial distance crosslinks would need to reach if telopeptide is extended (blue, see also A). Red bar shows short axial distance spanned by the folded telopeptide (green, see also E). Monomers are approximately 1.3 nm distant from each other in lateral space (B). Hydroxylysine is defined as single letter code “U”. A and C is a molecular surface rendering to show the model outline (see methods), not electron density.

**Fig. 5.** Cross-species comparison of TEM images of collagen type I (left) and type II (right) fibrils. Human quadriceps tendon (A), rat-tail tendon (B), lamprey notochord (C), Bovine articular cartilage (D), human articular cartilage (F).
**Supplementary Table 1.**
N-telopeptide secondary structure prediction. The turns for the electron density (Fig. 2 and S. Fig. 3) and real-space model (Fig. 3) are highlighted in green.

**Supplementary Table 2.**
C-telopeptide secondary structure prediction. The turns for the electron density and real-space models are highlighted in green. The turns for the electron density (Fig. 2 and S. Fig 3) and real-space model (Fig. 4) are highlighted in green.

**Supplementary Fig. 1.**
Hierarchy of the structure of fibrillar collagen. Note, that the specific structure of the type II microfibril has not been verified, here we indicate only that it could be similar to the type I collagen microfibril. Nevertheless the axial organization of the microfibril fits the Hodge-Petruska scheme. Red stick bonds show the location of the Gly-Xaa-Yaa, Gly in the super-helix interior.

**Supplementary Fig. 2.**
Composite profile of the Log of integrated intensities for 35 meridional orders of native (black), and iodine (gray) and platinum (white) derivatives of collagen type II fibrils from lamprey notochord.

**Supplementary Fig. 3.**
Electron density data, models and determination of gap / overlap ratio.
- a) Electron density models with extended telopeptides (blue), and folded (red). The model with the folded telopeptides was converted to the same resolution (1.9nm) as the experimental electron density through Fourier and reverse Fourier transforms and is shown in Fig. 2.
- b) Determination of Gap / Overlap ratio. The gap / overlap ratio of the one dimensional electron density map is largely defined by the first few meridional orders. The first order in particular dominates, with its large tail that tends to equalize the gap and overlap. This may be countered somewhat by reducing the size of the first order (ten times, in contrast to Fig. 2 where the first order is unaltered). This leaves enough signal to help define the gap and overlap and reduces (but does not completely remove) the tail of the first order in the overlap. The tyrosine (Y) residues found at the ends of the telopeptides (Y6 and Y1058) are found to be shifted towards the center of the overlap region (iodine difference Fourier and Patterson maps, Fig. 2, positions marked here with yellow lines). The amount by which they are shifted is too much to be accounted for by chain compression and is most likely due to the folding of the telopeptides (Figs 3 and 4, see results and discussion). Structure prediction from the telopeptide amino acid sequence also supports this observation (S. Tables 1 and 2). Horizontal arrows represent the distance between the marked heavy atom positions (yellow lines) and the turn position, whilst vertical red lines mark this extent on the map. The discrete (non-smoothed model electron density map with folded telopeptides, S. Fig. 3a) is shown with a red dashed line. The tail of the two highest peaks corresponds to the turn of the folded telopeptides and it is seen that these intersect at the calculated position of the folds of the telopeptides at the 60% of overlap region height line. This we feel best measures the extent of the overlap region (0.42 D). Lastly, the mid-line (cyan) marks the equidistant point between 60% height of overlap and 60% depth of gap region. This intersects the native density map at a value that well matches the expected position of the end of the C-terminal telopeptide fold but is 0.015D removed from the expected position of the N-telopeptide fold. However, this discrepancy is within the expected margin of error of the study (0.03D).

**Supplementary Fig. 4.**
Micrographs of positively stained sections of rat-tail tendon (A), lamprey notochord (B), human articular cartilage (C), bovine articular cartilage (D), human quadriceps tendon (E), and collagen fibrils (type II fibril bundle / thick fibril lower, type II ‘thin’ fibril upper) in suspension, isolated from human articular cartilage (F).
Fig. 2
Fig. 3
Fig. 4
Fig. 5
In situ D-periodic molecular structure of type II collagen
Olga Antipova and Joseph P.R.O. Orgel

J. Biol. Chem. published online January 6, 2010

Access the most updated version of this article at doi: 10.1074/jbc.M109.060400

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

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

Supplemental material:
http://www.jbc.org/content/suppl/2010/01/06/M109.060400.DC1