Collagen XI is a heterotrimeric molecule consisting of α1, α2, and α3 collagen chains (6). The α3(XI) chain is an overglycosylated form of the α1(II) collagen chain (7), whereas the α1(XI) and α2(XI) chains are distinct gene products (8). The α1(XI) and α2(XI) chains contain similar large, N-terminal domains (NTDs), comprising a "PARP" or "PARP-like" domain, a variable region, and a minor triple helix (Fig. 1; Refs. 9 and 10). The PARP (proline-arginine rich protein) domain was originally isolated as a distinct protein from bovine cartilage (11) but was later demonstrated to be a fragment of the N terminus of the α2(XI) chain (12). Recently, the proteolytic processing site of the α1(XI) chain has been identified (13, 14), and since nearly all of the domain is removed by this processing event, both the PARP-like and PARP domains are now termed the amino-propeptides (Npp). The Npp domains are structurally homologous to modules found in seven other collagen chains, as well as the glycoproteins laminin and thrombospondin (15, 16).

Both the α1 and α2 chains contain a variable region that lies between the Npp and the minor helix, and alternative splicing of the mRNA within these variable domains generates considerable sequence diversity (10, 17, 18). During chondrocyte differentiation, splicing of the α2(XI) mRNA rapidly converges to expression of a single protein isoform that lacks all segments encoded by alternatively spliced exons (19). Conversely, alternative splicing of the α1(XI) mRNA at two separate sites during the differentiation process results in divergence from a single species to a complex set of protein isoforms. Six possible isoforms are synthesized, which are named according to the peptide encoded by the exon: p6a, p6b, p8, p[6a+b]8, p[6b+8] and p0, an isoform that contains constitutively expressed regions only (Fig. 1a).2 Alternative splicing leads to considerable structural complexity, since the p6a and p8 peptides are acidic while the p6b peptide is extremely basic (18).

Although collagen XI comprises a relatively small amount of the total collagen present in cartilage fibrils, its presence is essential for the regulated assembly, organization, and development of the cartilage. This is evidenced by studies on the
autosomal recessive chondrodysplasia (Cho) mouse, where a frameshift in the α1(XI) gene leads to a functional knock-out of the gene product (20). The cartilage of these mice is atypical in that the collagen fibrils are much thicker than normal (21, 22), an observation that is consistent with a regulatory role for type XI collagen in fibril assembly. It has been proposed that the mechanism by which this occurs is similar to that by which collagen V limits the diameter of type IV fibrils in the cornea (23, 24). Here, the NTD of type V collagen at the fibril surface inhibits further accretion of collagen I monomers by steric hindrance or by electrostatic mechanisms (25). In accordance with this model, the NTD of type XI collagen would limit the diameter of heterotypic cartilage fibrils and this may rely in part on the shape and dimensions of the NTD. The goal of this study, therefore, was to examine the structure of the Npp alone or with the alternatively spliced peptides, p8 and p[6α+8] to further elucidate the regulatory mechanism. These two isoforms are representative of both the non-cartilage form and up to 30% of the α1(XI) in cartilage (26).

**MATERIALS AND METHODS**

**Construction of the Mammalian pCEPD47-p8 and pCEPD47-p[6a+8] Expression Vectors**—DNA constructs that code for the p8 and p[6α+8] isoforms were generated by reverse transcription-PCR using RNA from an immortalized rat chondrosarcoma cell line as a template (27). To amplify the sequence, the upstream primer used was 5′-ATT AGC TAG CTG CTC CAG TTT ATA TAG TG-3′, and the downstream primer was 5′-ATT ACT CGA GAC TAG TGA TGG TGA TGG TGG GGT TCA ACA GCA GGT TCC CC-3′ (Fig. 1). The upstream primer included a unique Nhel site, while the downstream primer encoded six histidine codons in tandem, a stop codon, a SpeI site, and a unique XhoI site, at the 3′ end. These primers were based on published cDNA sequence of the rat α1(XI) NTD (18). The sequences were amplified using the Advantage HF PCR kit (CLONTECH) with the following amplification conditions: an initial denaturation at 94 °C for 1 min, 30 cycles of denaturation (94 °C for 45 s), annealing (65 °C for 30 s), elongation (72 °C for 1 min), and a final elongation for 10 min at 72 °C. The PCR products were analyzed on a 2.5% Nusieve agarose gel (3:1; FMC Bioproducts) and were cloned into the Nhel-XhoI-cut pCEP4 mammalian episomal expression vector (Invitrogen), which had been modified to include the signal peptide and 5′-untranslated region of the basement membrane protein BM40 (28). Secreted recombinant proteins retain four amino acids from this signal peptide at the N terminus. Positive clones were selected by amplicolin resistance in *Escherichia coli* and their identity verified by DNA sequencing. The plasmid was harvested and transfected into human embryonic kidney (HEK) 293 cells that express the EBNA-1 protein (Invitrogen) using Pfx-2 lipid (Invitrogen). Cells were grown in DMEM/F-12 containing 10% fetal calf serum and 1% penicillin/streptomycin (Gibco-BRL). Cells were passaged and selected in G418 (400 μg/ml) and G418 (50 μg/ml). These cells were expanded for large scale culture in serum-free DMEM/F-12. Expression of recombinant protein was determined for up to 2 months, and medium was collected twice a week. Pooled media was clarified by centrifugation before proceeding with protein purification.

**Constitution of the Prokaryotic α1-Npp Recombinant Vector**—The α1-Npp sequence was amplified with the upstream primer 5′-TAA CTA GCT CCA GGT GAT ATA CTG AAA GAT TTA GAT TTT-3′ (Fig. 1), which included a unique Nhel restriction site. The downstream primer was 5′-ATT AGG ATG AAC CAC TAC TGA TGG TGA TGG TGG TGG GGT TCA ACA GCA GGT TCC CC-3′, which encoded six histidine residues, a stop codon, a SpeI site, and a unique BanHI site at the 3′ end. The α1-Npp sequence was amplified using the Advantage HF PCR kit (CLONTECH), cloned into the pET 11a expression vector (Stratagene), and the identity was verified by sequencing. Positive clones were selected and plasmid DNA was used to transform BL21 (DE3) Flys-S competent cells, which express T7 polymerase (Invitrogen). These were grown in LB medium, and α1-Npp was purified from the inclusion bodies of these cells by solubilizing in 8 M urea, 14 mM hydroxylamine, and 1 mM benzamidine. The protein solution was dialyzed against 20 mM sodium phosphate, pH 7.8, containing 0.25 mM NaCl and then dialyzed against chelating column start buffer (see below).

**Purification of Recombinant Proteins**—Recombinant p8 and p[6α+8] proteins were precipitated from serum-free medium by 30% ammonium sulfate and resuspended in start buffer (1 mM NaCl, 20 mM sodium phosphate, 2 mM urea, pH 7.1) containing protease inhibitors (1 mM N-ethylmaleimide, 1 mM [4-(2-aminomethyl)benzenesulfonyl]fluoride, hydrochloride and 1 mM benzamidine). Insoluble material was pelleted by centrifugation (18,000 × g for 30 min) and the resultant supernatant filtered through a 0.8-μm filter. This was loaded onto a 5-ml HiTrap chelating affinity column (Amersham Pharmacia Biotech) charged with 0.5 M sodium chloride and eluted with Identical buffers. The absorbance was monitored continuously at 280 nm, and the flow rate was 2 ml/min. Bound proteins were eluted by a step gradient to 0.25 M imidazole. Proteins in each fraction were analyzed by SDS-PAGE (7.5% gel, reducing conditions) and stained with Coomassie Blue. Fractions containing recombinant protein were pooled, concentrated to 1 ml by ultrafiltration (Amicon) using a YM-30 filter, and passed over a Superose 12 column (Amersham Pharmacia Biotech) equilibrated in 0.1 M NaCl, 20 mM sodium phosphate, 2 mM urea, pH 7.2. The absorbance was monitored continuously at 280 nm, and the flow rate was 0.25 ml/min. Purified recombinant protein was then concentrated on a Centricon filter unit (Amicon). α2-Npp was extracted with guanidine HCl from bovine nasal cartilage and purified as described previously (11).

**Analysis of Folding by Disulfide Mapping—**Purified p8 and p[6α+8] recombinant proteins were digested with CNBr in 70% formic acid to isolate the N terminus containing the disulfide-bonded cysteines. After digestion the sample was subjected to reverse-phase HPLC (Vydac C18 column 218TPF52), and the N-terminal peptide was identified by Edman sequencing (Routine 3.5 chemistry, ABI Sequencer). This peptide was digested with V8 protease (Roche Molecular Biochemicals) in 50 mM NH4OAc, pH 6.8 at 20 °C for 18 h, then purified by reverse-phase HPLC with the following conditions: solvent A, 0.1% trifluoroacetic acid/H2O; solvent B, 0.08% trifluoroacetic acid/acetonitrile; gradient, 0–1.6% B, 63–29.6% B, 95–60% B, 105–80% B, 105.1–1.6% B. Disulfide-bonded peptides were identified by Edman sequencing.

One hundred pmol of α1-Npp protein was digested with 2 pmol of sequencing-grade trypsin (Promega, Madison, WI) for 4 h in 50 mM ammonium bicarbonate. 1 pmol of the peptide digest was pressure-loaded onto a 75-μm capillary HPLC column for on-line analysis by liquid chromatography–tandem mass spectrometry (LC-MS/MS) in an ion-trap mass spectrometer (LCQ, Finnigan, San Jose, CA) as described (29). As peptides eluted from the column, their mass/charge ratios were measured in the mass spectrometer. A second stage of mass spectrometry was then performed automatically, where a selected peptide ion was fragmented to collect sequence information. Disulfide bonds are stable under the conditions used, and no bond breakage between cysteines occurred. A mass/charge ratio 1315.8, [M + 3H]+ corresponding to a trio of disulfide-bonded peptides was encountered during the run. A tandem mass spectrum was automatically collected, which was used to determine the correct order of disulfide arrangement. This trio was the only peak not present in an identical run where the peptide digest was reduced with 10 mM dithiothreitol for 1 h at 37 °C.

**Glycosylation Analysis—**ScanProsite was used to identify potential N-glycosylation sites, while the NetOGlyc 2.0 prediction server was used to predict potential GalNAc O-glycosylation sites of α1(XI) NTD. To assess whether the recombinant p[6α+8] protein was glycosylated, a sample was dialyzed against 20 mM sodium phosphate, 2 mM urea, pH 7.1. lyophilized, and incubated overnight at 37 °C with either 0.5 units of N-glycosidase F alone or in combination with 0.25 milliunits of O-glycosidase (Roche Molecular Biochemicals). Digested and undigested control samples were separated on a 7.5% SDS-PAGE gel and stained with Coomassie Blue.

**Molecular Sieve Chromatography—**For analysis of protein purity, recombinant p[6α+8] recombinant protein was passed over a Superose 12 molecular sieve column. This was performed in both denaturing (8 M urea, 14 mM β-mercaptoethanol, 0.1 M NaCl, 0.1 M Tris-HCl, pH 7.1, 0.1% Triton X-100) and non-denaturing conditions (0.1 M NaCl, 0.1 M Tris-HCl, pH 7.1). The absorbance was monitored continuously at 280 nm, and the flow rate was 0.25 ml/min. SDS-PAGE was used to analyze every second fraction of the denatured buffer run to determine the elution position of p[6α+8]. The elution position was compared with a set of known globular proteins (670 to 1.35 kDa (Bio-Rad); 67 to 13.7 kDa (Amersham Pharmacia Biotech)) resuspended in identical buffers and run immediately prior to each of the experimental runs.

**Rotary Shadowing—**The purified isoforms (10 μg/ml) were diazylated against 0.2 M ammonium bicarbonate, sprayed onto mica, and shadowed with platinum (50%) and carbon (50%) in a high-voltage electron microscope (Hitachi), using a carbon support. The resulting replicas were viewed in transmission electron microscopy. Measurements were estimated from photomicrographs using a calibrated eyepiece. Only molecules with defined, extended tails were measured. To make the measurements more accurate, the size of the platinum particles, measured from the background, was subtracted from this value. Statistical analysis was performed and a standard error of the mean calculated.

Downloaded from http://www.jbc.org/ by guest on July 25, 2018
**Circular Dichroism**—Purified recombinant isoforms (0.5–1 mg/ml) were dialyzed extensively against 0.1 M sodium phosphate, pH 7.2. Each sample was analyzed in the far-UV range (260–180 nm) at 24 °C by CD spectroscopy on a Jasco J-500 spectropolarimeter interfaced to a computer, using a quartz cell with a path length of 100 μm. Spectra were smoothed, and buffer base lines were subtracted. The secondary structure was subsequently estimated using the variable selection method with an initial set of 33 basis spectra (31). This method of analysis is thought to be advantageous in estimating the β-sheet content of a protein (32). The concentration of each sample was determined in triplicate by amino acid analysis. Samples were gas-phase hydrolyzed using 300 μl of 6 N HCl, 2% phenol at 110 °C for 22 h, and the hydrolysates were analyzed on a Beckman 6300 amino acid analyzer using sodium citrate buffers. For determination of the structure of the variable region, the difference between the normalized spectra of the α1-Npp and the p6a+8 domain was obtained using a calculation that accounted for the different number of amino acids present in each domain. The resultant spectrum of the variable domain was then solved.

**Secondary Structure Predictions**—The secondary structure of the α1-Npp and variable region of the α1(XI) NTD were predicted from the following programs: (a) information theory with nil decision constants (Gor1; Ref. 33), (b) Gor2, a modification of Gor1 (34), (c) Gor 4, a further improved version of Gor1 (35), (d) self-optimized method for protein secondary structure prediction (SOPM; Ref. 36), (e) SOPMA, an improved version of SOPM (37), (f) PHDsec (38), (g) hierarchical neural network (39), (h) Predator (40), and (i) I-PrEDict (41). These prediction programs can all be accessed at the EXPASY web server, where the default options, if any, for e.g. window lengths and similarity thresholds, were chosen. Not all these programs predicted the same secondary structure for each sequence, so the prediction was based on agreement of at least four out of the nine methods implemented.

**RESULTS**

**Purification of Recombinant Protein**—The α1-Npp domain alone or the complete p[6a+8] isoform (Fig. 1b) were expressed as recombinant proteins in prokaryotic and eukaryotic systems, respectively, as sufficient quantities could not be purified from cartilage. Recombinant p[6a+8] was initially purified by chelating affinity chromatography utilizing the 6-histidine tag incorporated into the construct (Fig. 2a). SDS-PAGE analysis and western blotting with monoclonal antibodies directed against the 6-histidine tag indicated that the material in pool B contained the recombinant protein (results not shown). However, Coomassie Blue staining of the same material showed the presence of a high M₈ contaminating protein (Fig. 2c, lane 2). The fractions containing partially purified recombinant protein were pooled, concentrated and passed over a Superose 12 molecular sieve column (Fig. 2b). SDS-PAGE analysis and Coomassie Blue staining of the peak fractions indicated that the high M₈ contaminating protein was successfully removed (Fig. 2c, lane 3), leaving substantially pure recombinant p[6a+8] protein in the second peak (Fig. 2c, lane 4). These fractions were pooled and the protein concentrated for use in structural analyses. Recombinant α1-Npp was purified in an identical manner and the purified product is shown (Fig. 2c, lane 5). The typical yields of purified recombinant protein were 1 mg of p[6a+8]/4 liters of DMEM/F-12 medium or 1 mg of α1-Npp/liter of LB medium.

**Analysis of Folding by Disulfide Mapping**—The disulfide bonds formed by the recombinant p[6a+8] protein were analyzed by peptide mapping under non-reducing conditions. After digestion with cyanogen bromide and chymotrypsin, Edman sequence analysis of disulfide-linked peptides revealed that Cys-1 was linked to Cys-4 and that Cys-2 was linked to Cys-3 (Fig. 3a). There remains a possibility that a small amount of protein was not folded correctly but as the protein migrated as a single species on SDS-PAGE and eluted as a single, symmetrical peak on molecular sieve chromatography, it is likely that most, if not all, of the protein was folded correctly.

The proper disulfide bond pairing of the recombinant α1-Npp was confirmed by LC-MS/MS. Disulfide analysis by LC-MS/MS of a tryptic digest of α1-Npp resulted in the detection of the mass/charge ratio for a trio of predicted disulfide cross-linked peptides with two possible cross-linked conformations (Fig. 3b). A tandem mass spectrum corresponding to this peptide revealed the correct disulfide order because disulfides did not break under the conditions employed and amino acid sequence could be verified across the S-S bonds. Only a single peak disappeared in a comparison between native and reduced tryptic digests, and no other masses corresponding to any other predicted disulfide-linked peptides were identified, suggesting that the conformation detected was the only one present in the recombinant protein. The disulfide bond pattern detected for both the eukaryotic and prokaryotic recombinant proteins was identical to that of the highly homologous α2-Npp domain isolated from cartilage (11), indicating that the recombinant proteins are likely to be correctly folded.
Prosite was used to analyze Np1.

Blue staining. Fractions from peak 2 were pooled and concentrated. peak 4 lane 4, purified p6a cobalt chelating column; lane 3, dashed line an imidazole step gradient (profile of the cobalt chelating column. Bound proteins were eluted with tion profile of the Superose 12 molecular sieve column. His-tagged recombinant protein; C with 8 glycosylation analysis performed to map the disulfide bonds, the asparagine residue of the amino acid sequence -NISE- was identified as a potential site within the p8 peptide. To assess whether the recombinant p8 protein, and in both cases double digested with N- and O-glycosidase did not further increase the rate of migration (results not shown). Moreover, during the amino acid sequence analysis performed to map the disulfide bonds, the asparagine residue of the amino acid sequence -NISE- was absent, consistent with an N-glycosylation at this site within the p8 peptide.

Molecular Sieve Analysis of Eukaryotic Recombinant p[6a+8]—Following the molecular sieve purification of the p[6a+8] isofrom, standard globular proteins were passed over the column in identical buffer conditions. It was observed that, in non-denaturing buffer, the p[6a+8] eluted at a much greater apparent molecular mass (~280 kDa) than its mass calculated from its amino acid sequence (Fig. 5a). The elution of p[6a+8] as a single, sharp, symmetric and included peak in non-denaturing conditions was an indication that the greater apparent molecular mass was not due to multimer formation, but rather an asymmetric conformation. When chromatographed under denaturing conditions (8 M urea, 14 mM β-mercaptoethanol, 0.1% Triton X-100; Fig. 5b), the p[6a+8] eluted at a position corresponding to approximately 66 kDa, the same apparent molecular mass detected by SDS-PAGE. Moreover, the α1-Npp domain eluted at a position corresponding to its molecular mass in non-denaturing conditions (results not shown), suggesting that asymmetry resides in the variable region.

Rotary Shadowing—The indication of an asymmetric conformation of p[6a+8] was investigated further by use of rotary shadowing and electron microscopy. Rotary-shadowed images of both the recombinant α1-Npp and the α2-Npp domain showed a homogeneous population of small round particles, indicating formation of a compact, globular domain (Fig. 6, a and b). This observation is similar to the rotary shadowing images of the NC11 domain, a homologous domain found in type XVI collagen (42). Conversely, the p[6a+8] isofrom, which contains variable region in addition to the Npp domain, displayed two distinct folding domains: a globular “head” and an extended “tail” (Fig. 6c). Measurement of these images revealed that the α1-Npp and α2-Npp domains were similar in size, 8.4 (± 0.17) nm and 10 (± 0.16) nm in diameter, respectively (Fig. 6, a and b). These measurements are only approximate since the size of the Npp domain approaches the limits of resolution of this technique. The extended tail of the p[6a+8] isofrom was approximately 16 (± 0.4) nm long (Fig. 6c). Rotary shadowing of p8 and p6b recombinant proteins also displayed a globular head and an extended tail (results not shown). It is possible that the rotary shadowing technique may exaggerate the appearance of extension and rigidity. However, molecular sieve chromatography indicated a highly asymmetric conformation in solution so that the variable region, although probably somewhat flexible, is quite extended on average.

Mass/Length Determination—The mass/length ratio of the variable region was determined from rotary-shadowed images and was found to be in the range of 1050–1530 Da/nm. When compared with other extended protein structures, the mass/length ratio of the variable region is similar to a two-stranded coiled-coil (1420 Da/nm; Ref. 43) and a collagen triple helix (1000 Da/nm), two of the most extended polypeptide structures. It should be noted, however, that in each of these comparisons more than one peptide chain contributes to the overall structure. Our results suggest that an equally extended structure can be achieved by a single polypeptide without benefit of significant amounts of periodic secondary structure such as α-helix and β-sheet (see below). We cannot exclude the possibility that, in vivo, additional structure results from an interaction between this region and the variable domain of the α2(XI) chain, although this seems unlikely as the α2(XI) protein isoform predominantly expressed in cartilage lacks the alternatively spliced variable region (19).

Circular Dichroism—Experimental evidence for the secondary structure of the recombinant α1-Npp and p[6a+8] proteins was obtained by CD spectroscopy (Fig. 7). The α1-Npp spectra
Fig. 3. Disulfide mapping of p[6a+8] and α1-Npp. a, results of Edman sequence analysis of p[6a+8] revealed that Cys-1 is linked to Cys-4 and that Cys-2 is linked to Cys-3. A dash (−) denotes a blank in the sequence indicating the first half of a cysteine pair, and an asterisk (*) denotes a cysteine residue. A schematic representation of the primary structure of the α1-Npp domain shows the enzymatic digestion sites and the disulfide bond arrangement. Solid lines represent disulfide bonds, rectangles depict CmBr cleavage sites, and diamonds represent chymotrypsin cleavage sites. b, tandem mass spectrum of a trio of disulfide-bonded peptides with a precursor mass/charge ratio 1315.8 [M-H]−. The two potential disulfide bonding scenarios are shown. This large ion was fragmented on average at a single peptide bond along the chain. The two resulting fragment (b and y) ions were recorded in the tandem mass spectrum and were later reconciled with the sequence. Because disulfide bonds are not broken here, the peptide sequence can be followed across the S-S bond, and the correct bond orientation determined. Arrows represent peptide bond cleavage between the two cysteines of the middle peptide producing b and/or y ions, which unambiguously determined disulfide bond assignment.

Fig. 4. Glycosylation analysis of p[6a+8]. Recombinant p[6a+8] protein was separated by SDS-PAGE (7.5% gel, reducing conditions), with (+) or without (−) prior treatment with N-glycosidase F. Coomassie Blue staining showed that the treated sample migrated faster indicating the removal of N-linked glycan chains. Molecular size markers are shown (kDa).

showed negative ellipticity at 216 nm and a peak of positive ellipticity at 198 nm, which together are indicative of β-sheet structure (Fig. 7a; Refs. 44 and 45). The lack of negative minima at 208 and 222 nm and positive ellipticity at 192 nm suggested that this protein contained relatively little α-helix. The α2-Npp spectrum was slightly different in that the curve was broader between 216 and 208 nm, and the positive ellipticity peak had shifted from 198 to 193 nm (Fig. 7b). The results of this analysis were in good agreement with secondary structure predictions of the variable region, although somewhat more β-structure was detected than predicted (Table I). A similar difference analysis of the p[6a+8] and p8 spectra indicated that the p6a component contained mainly β-turn (results not shown).
FIG. 5. Molecular sieve analysis of p(6a + 8). Pure p(6a + 8) was passed down a Superose 12 gel filtration column either in non-denaturing conditions (a) or in buffer containing 8 M urea, 14 mM captoethanol, and 0.1% Triton X-100 (b). As the detergent interferes with absorbance at 280 nm, the elution position of p(6a + 8) was determined by running every second fraction on SDS-PAGE. The molecular sizes of globular standard proteins in both buffers were plotted as a function of their elution volume, and the apparent molecular mass of p(6a + 8) was calculated from the standard curves. In non-denatured buffer (a), p(6a + 8) was found to elute at ~280 kDa, whereas, under denaturing conditions (b), it eluted at ~66 kDa, which is consistent with the position it migrates on SDS-PAGE. The anomalous migration of p(6a + 8) under non-denatured conditions is indicative of an asymmetric conformation, rather than aggregation or multimerization. It should be noted that, although it looks like the protein elutes at the same position, this is a chance occurrence as urea changes the sieving capacity of the column.

DISCUSSION

The model for the regulation of fibrillogenesis by type XI collagen is related to that proposed for type V collagen, where the NTDs play the critical role in control of fibril diameter. For example, in the heterotypic collagen fibrils of the cornea, the triple helical domains of type V collagen are sequestered in the interior of the fibril, whereas the retained NTDs of the $\alpha$1(V) chains are excluded from the interior of the fibril, accumulate on the surface, and thus eventually sterically hinder the further deposition of type I collagen molecules onto the fibril (23–25). A similar mechanism has been proposed for type XI collagen and is supported by in vitro fibrillogenesis assays with cartilage collagens that demonstrate uniformly thin collagen type II fibrils will assemble only when collagen XI is present (46). In the following discussion, emphasis has been placed on the effects of $\alpha$1-NTD on fibrillogenesis because it is relatively long-lived in the tissue, compared with the $\alpha$2-Npp, which is rapidly processed after synthesis (47).

The major triple helical domain of type XI collagen is also sequestered within the collagen fibril of fetal cartilage (1, 3). However, it is clear from the molecular dimensions of the $\alpha$1-Npp, the variable region, and the minor triple helix that they would not be accommodated within the confines of the gap region of a fibril (1.5/2.5 × 32 nm; Fig. 5a; Ref. 46). Indeed, the $\alpha$1-Npp has been localized on the surface of collagen fibrils in fetal cartilage (48), and the same is true for the variable region peptides. It is not known, however, where the minor helix resides, whether it remains in the interior of the fibril or functions to extend the NTD to the fibril surface, so both possibilities are shown (Fig. 8b). In terms of length, the minor helix alone (~20 nm) would be sufficient to exclude the $\alpha$1-Npp from the interior of the 20-nm diameter fibrils, although the variable region (16 nm in the most extended form) would also suffice. The NTD may turn back on itself, as has been shown with the N-telopeptides of type I collagen (49), or it may adopt other orientations with respect to the fibril (Fig. 8c). All possibilities still allow the N-terminal region, especially the globular $\alpha$1-Npp domain, to limit the appositional growth of the fibril by blocking the further accretion of collagen monomers.

However, a mechanism of regulation of fibrillogenesis based solely on consideration of steric effect of the Npp may be simplistic. For example, despite their homology and similar molecular dimensions (including the Npp), 10% content of type XI collagen is sufficient to produce the 15–20-nm fibrils of cartilage (1), whereas 20–30% content of type V collagen is required to obtain the 25-nm fibrils of the corneal stroma (24). In addition, if fibril diameter is strictly governed by the steric effect of N-propeptide domains, then one would expect that collagen fibrils composed of pN-collagen to be thinner than normal. However, in patients with Ehlers-Danlos syndrome type VIIIB and in animals with dermatosparaxis, diseases where fibrils contain predominantly pN-collagen I molecules, normal diameter fibrils are found. Instead, their surfaces are highly convoluted as surface area is maximized to accommodate these extra propeptide domains (50, 51). Thus it is not likely that a steric effect alone is sufficient to explain the effects of retained N-propeptides on fibrillogenesis in these tissues. Other factors clearly influence fibrillogenesis in all of these systems, factors such as rate and order of propeptide processing of the major collagens, types I and II (52), and the participation of fibril-associated proteins such as decorin, fibromodulin, and TRAMP (53–55).

The CD calculations presented here are consistent with previous secondary structure predictions, which indicate that the $\beta$-sheet of the $\alpha$1-Npp domain could be distributed into nine

---

3 Morris, N. P. Oxford, J. T., Davies, G. B. M., Smoody, B. F., and Keene, D. R. (2000) *J. Histochem. Cytochem.*, in press.
β-strands (15, 45). This region is structurally related to a domain of the glycoprotein thrombospondin that is termed the Tsp1 module (16). This module, which is compatible with an Ig-fold structure, has also been identified in other collagen chains, namely α1(V), α1(IX), α2(IX), α1(XI), α1(XIV), α1(XVI), and α1(XIX) (15, 16, 42). A superfamily of proteins

**FIG. 6.** Rotary-shadowed images of α1-Npp, α2-Npp, and p[6a+8]. Pure α1-Npp (a), α2-Npp (b), and p[6a+8] (c) were sprayed onto mica, shadowed with platinum, and visualized by transmission electron microscopy. Both the α1-Npp and α2-Npp formed a single compact folding domain (black arrowheads), whereas the p[6a+8] isoform showed two distinct folding domains, a globular “head” and an extended “tail” (white arrowheads). Insets show a higher magnification of each isoform. Bar = 20 nm. Measurements of the head and tail domains were measured from photomicrographs using a calibrated eyepiece, and the mean diameter or length was determined (mean ± standard error; n = 50).

**FIG. 7.** Far-UV CD spectra of α1-Npp, α2-Npp, and p[6a+8]. Purified proteins were dialyzed against 0.1 M sodium phosphate, pH 7.2, and analyzed in the far-UV range (260–180 nm) at 24 °C by CD spectroscopy. CD spectra of the α1-Npp domain (a), the α2-Npp domain (b), and the p[6a+8] isoform (c). All spectra were smoothed, and buffer base lines were subtracted. d, a difference spectrum was generated from the α1-Npp and p[6a+8] spectra to determine the structure of the variable region alone.
that contain Ig-like folds has thus emerged, members of which are related due to their structural similarity rather than their sequence identity. Members of this family have been shown to contain 7–10 β-strands and are classified into five different subtypes (56). Although the β-structure of the Npp domains could be arranged into nine β-strands, it remains unclear as to which subset of Ig-fold that they belong and this will only be resolved by x-ray crystallography. Nevertheless, the results presented have broad implications for the structure and function of this motif in other extracellular matrices.

The function of the variable region is unknown, but the data presented here suggest several possibilities. The location of the variable region on the fibril surface suggests that, in addition to facilitating the exclusion of the globular α1-Npp domain from the fibril core, it may also participate in interactions with other matrix components, thus greatly magnifying the steric effect or mitigating fibril fusion. Such an interaction could contribute to the biomechanical integrity of cartilage and is consistent with the loss of cohesion observed in the cartilage of the Cho mouse (21, 22). The variable region peptides are either acidic (p6a and p8) or basic (p6b; Ref. 18), which suggests a potential for differential interactions, and it should be noted that the p6b peptide contains the minimal amino acid sequence for heparin binding (b-Xn-b; Ref. 57). Alternatively, the identity of the variable region could influence the rate or extent of proteolytic removal of the α1-Npp. The proteolytic processing site is positioned seven amino acids distal to the variable domain at the sequence KAAQA|QEPHIDE (13, 14). Recent studies have implied that BMP-1 is responsible for processing as this enzyme cleaves the Npp from α1(V) at a homologous sequence (58). This enzyme is also responsible for cleavage of the C-propeptide from type I collagen, and molecular modeling shows that this region has a highly extended structure (59). This implies that BMP-1 requires an extended substrate structure for activity. Preliminary results suggest that the α1-Npp is cleaved more rapidly from an isoform containing an extended domain (p6b) than from p0, which contains no variable region.4 Interestingly, this does not seem to affect fibril diameter, as there was no difference between fibrils that contained the p6b isoform and those that did not.3 This suggests that the differential rate of processing may serve a different function. In addition, the extended conformation of the p[6a+8] isoform suggests that the alternatively spliced peptides of the variable region are organized linearly. This arrangement would readily accommodate the complex modulation of its primary structure by alternative splicing and would also facilitate potential alterations in the function of each individual isoform.

In conclusion, the structural organization of the NTD presented here, the differential rate and extent of processing of the α1-Npp and α2-Npp (47), and the developmentally regulated distribution of isoforms based on alternative splicing within the variable region4 together suggest that the NTD may contribute to functions of type XI collagen more diverse than regulation of fibrillogenesis.

**Acknowledgments**—We thank Doug Keene and Sara Tufa for the rotary shadowed images. We also acknowledge the expert technical assistance of Joe DeScala in constructing the α1-Npp recombinant vector and Bruce Donaldson for help with recombinant protein purification.

**REFERENCES**

1. Mendler, M., Eich-Bender, S. G., Vaughan, L., Winterhalter, K. H., and Bruckner, P. (1989) J. Cell Biol. 108, 191–197
2. Nah, H. D., Barembaum, M., and Upholt, W. B. (1992) J. Biol. Chem. 267, 22561–22566
3. Petit, B., Ronziere, M. C., Hartmann, D. J., and Herbage, D. (1993) Histochem-

---

4. J. T. Oxford, unpublished observations.
Structural Organization of Distinct Domains within the Non-collagenous N-terminal Region of Collagen Type XI
Kate E. Gregory, Julia T. Oxford, Yanwen Chen, Jay E. Gambee, Steven P. Gygi, Ruedi Aebersold, Peter J. Neame, Diane E. Mechling, Hans Peter Bächinger and Nicholas P. Morris

J. Biol. Chem. 2000, 275:11498-11506.
doi: 10.1074/jbc.275.15.11498

Access the most updated version of this article at http://www.jbc.org/content/275/15/11498

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

This article cites 57 references, 20 of which can be accessed free at http://www.jbc.org/content/275/15/11498.full.html#ref-list-1