Type V Collagen: Molecular Structure and Fibrillar Organization of the Chicken α1(V) NH₂-terminal Domain, a Putative Regulator of Corneal Fibrillogenesis

Thomas F. Linsenmayer,* Eileen Gibney,* Frank Igoe,* Marion K. Gordon,* John M. Fitch,* Liselotte I. Fessler,‡ and David E. Birk*

*Department of Anatomy and Cellular Biology, Tufts University Medical School, Boston, Massachusetts 02111; ‡Molecular Biology Institute and Department of Biology, University of California at Los Angeles, Los Angeles, California 90024

Abstract. Previous work from our laboratories has demonstrated that: (a) the striated collagen fibrils of the corneal stroma are heterotypic structures composed of type V collagen molecules coassembled along with those of type I collagen, (b) the high content of type V collagen within the corneal collagen fibrils is one factor responsible for the small, uniform fibrillar diameter (25 nm) characteristic of this tissue, (c) the completely processed form of type V collagen found within tissues retains a large noncollagenous region, termed the NH₂-terminal domain, at the amino end of its α1 chain, and (d) the NH₂-terminal domain may contain at least some of the information for the observed regulation of fibril diameters. In the present investigation we have employed polyclonal antibodies against the retained NH₂-terminal domain of the α1(V) chain for immunohistochemical studies of embryonic avian corneas and for immunoscreening a chicken cDNA library. When combined with cDNA sequencing and molecular rotary shadowing, these approaches provide information on the molecular structure of the retained NH₂-terminal domain as well as how this domain might function in the regulation of fibrillar structure.

In immunofluorescence and immunoelectron micros-copy analyses, the antibodies against the NH₂-terminal domain react with type V molecules present within mature heterotypic fibrils of the corneal stroma. Thus, epitopes within at least a portion of this domain are exposed on the fibril surface. This is in marked contrast to mAbs which we have previously characterized as being directed against epitopes located in the major triple helical domain of the type V molecule. The helical epitopes recognized by these antibodies are antigenically masked on type V molecules that have been assembled into fibrils.

Sequencing of the isolated cDNA clones has provided the conceptual amino acid sequence of the entire amino end of the α1(V) procollagen chain. The sequence shows the location of what appear to be potential propeptidase cleavage sites. One of these, if preferentially used during processing of the type V procollagen molecule, can provide an explanation for the retention of the NH₂-terminal domain in the completely processed molecule. The sequencing data also suggest that the NH₂-terminal domain consists of several regions, providing a structure which fits well with that of the completely processed type V molecule as visualized by rotary shadowing.

THE major component of the mature corneal stroma is a striated collagen fibril characterized by its small, uniform diameter (25 μm). Because these fibrillar properties are thought to be required for corneal transparency, elucidating the molecular mechanism(s) by which they are controlled is paramount to our understanding of the development and growth of a functional cornea, as well as of fibrillogenesis in general.

Striated collagen fibrils are composed largely of molecules of the fibrillar collagen class (31) arranged in a quarter-stagger array (19, 22). This molecular arrangement determines certain properties of a fibril. In itself, however, this arrangement can not be responsible for regulating fibril diameter, because it is the structural form of all striated fibrils, regardless of their diameter.

A number of factors have been suggested as being important in the regulation of fibril diameter. These include glycosylation of collagen molecules (26), procollagen processing (27, 41) and collagen-proteoglycan interactions (43). Recently, we have obtained evidence for the involvement of

a different type of regulation: the interaction of two or more different types of fibrillar collagen molecules, resulting in the assembly of what we have termed the heterotypic fibril (28, 29, 38, 47).

Evidence for such an arrangement came initially from the observation that epitopes for mAbs directed against the triple helical domain of type V collagen were antigenically masked in sections of corneal stroma, as well as in a number of other tissues (33, 34). Unmasking could be effected only by disrupting fibril structure (3, 15, 34) or by enzymatically removing type I collagen with a vertebrate collagenase selective for type I collagen (13, 15). This suggested that collagen types I and V were coassembled in such a way that the type I molecules rendered epitopes on the type V molecules unavailable for antibody binding. Double-label immunoelectron microscopy provided a definitive demonstration of such a heterotypic arrangement for these two collagen types (3). In addition, because the label for each collagen type occurred with the same 65–70-nm repeat predicted by the quarter-stagger model of fibril structure, these data suggested that an interaction existed between the two molecular types resulting in their coassembly. Other heterotypic combinations of collagens have since been observed in cartilage (38), skin and tendon (28, 29).

We hypothesized that the coassembly of different collagens could dictate certain properties of fibrils, such as diameter (32). The structural similarities of the fibrillar collagens would allow them to coassemble; their differences could sterically influence fibrillogenesis. We also hypothesized that the high concentration of type V collagen within cornea (15–20% compared to <5% in other tissues [37, 46]) could be responsible for the small uniform diameter of the stromal fibrils. Our studies employing in vitro fibrillogenesis (2) and corneal fibroblast cell culture (7) support this hypothesis. When both types of collagen molecules are copolymerized, increasing the proportion of type V collagen progressively decreases the diameter of the heterotypic fibril produced.

The native, completely-processed type V collagen molecule, as extracted from tissues, retains a large pepsin-sensitive domain at the NH$_2$-terminal end of its α1(V) chain (1, 4, 8, 11, 30). We will refer to this retained portion as the NH$_2$-terminal domain.[1] Most other fibrillar collagens, after processing from their procollagen forms, retain only short (~20 amino acid) peptides at their ends. The in vitro heterotypic fibrillogenesis experiments demonstrated that the fibril diameter-regulating properties of the type V collagen molecule reside largely within this NH$_2$-terminal domain (2). Molecules from which this domain had been artificially removed using pepsin digestion showed little effect compared to native type V molecules. Thus, the NH$_2$-terminal domain is a potential regulatory element.

Others have suggested from in vivo (16, 17) and in vitro studies (41) that the removal of propeptides from procollagen types I and III can influence fibrillogenesis. Because the

rates at which the propeptides are removed can vary for different collagens and for different tissues, this may be one factor in fibril growth. It may be through a similar type of steric mechanism that the NH$_2$-terminal domain of the α1(V) chain limits fibril diameter. However, because this domain is not removed during processing of the type V procollagen molecules, it theoretically can maintain a constant fibril diameter for an indefinite period of time. Recently, Chapman (5) has presented a formal model that generalizes these ideas.

In the present study, in order to examine further the properties of type V collagen within corneal collagen fibrils, and to provide a basis for further studies of the molecule, polyclonal antibodies have been produced which react with the NH$_2$-terminal domain, the putative regulatory domain within the α1(V) chain. These have been used for immunohistochemical examination of embryonic avian corneas, and for immunoscreening a chicken cDNA library. The immunohistochemical analyses showed that epitopes within this domain are available for antibody binding, even when the molecule is present within the heterotypic fibrils of the corneal stroma. This confirms that the NH$_2$-terminal domain does remain associated with the molecule after its assembly into fibrils, and that at least a portion of it is exposed at the fibril surface.

Sequencing of the cDNA clones has provided the conceptual amino acid sequence of the entire amino end of the chicken α1(V) chain. The chicken sequence has a number of similarities to the recently published human sequence (18, 45), but also a number of differences. Analysis of the sequence data suggests the presence of several separate regions within the amino-terminal end and the location of a putative propeptidase cleavage site which, if used in removal of the propeptide portion of the amino end, provides an explanation as to why the NH$_2$-terminal domain remains as a part of the completely processed type V collagen molecule.

Materials and Methods

Antibody Production and Characterization

---

[1]. The amino end of the pro-α1(V) chain begins at the initiation methionine and ends at the start of the major triple helix. It appears to be comprised of three separate domains which will be referred to as the signal peptide, the NH$_2$-propeptide (removed during procollagen processing), and the NH$_2$-terminal domain (retained with completely-processed type V collagen molecules). The three domains together will be referred to as the "entire amino end" of the α1(V) chain.

from these forms of the $\alpha1(V)$ chain by purified bacterial collagenase digestion and immunoprecipitated the partially and completely processed forms of type V. They did not show reactivity with any of the forms of the closely related $\alpha1'(V)$ chain (The $\alpha1'(V)$ chain is thought to be very similar to the $\alpha1(V)$ chain (12). It is found in some tissues, but if it is present in cornea, it is in very small amounts [unpublished observations].) nor with the non-collagenous peptides derived from this chain. They also failed to react with the major triple helical domain of chains from which the noncollagenous peptides had been removed by limited pepsin digestion.

Immunohistochemical Procedures

Rotary shadow electron microscopy of antibody binding to the completely processed, tissue form of type V collagen was performed as previously described (2, 36). Immunofluorescence and immunoelectron microscopy of corneas were also performed as previously described (3). For immunoelectron microscopy, both ultracryomicrotomy and pre-embedding were employed. The anti-NH$_2$-$\alpha1(V)$ antibodies were used at dilutions of 1:150 or 1:300; the mAbs were used at 10–25 μg/ml.

Molecular Biology

The anti-NH$_2$-$\alpha1(V)$ antibodies were utilized for immunoscreening a 10-d chicken embryo cDNA library (Clonetech Laboratories, Inc., Palo Alto, CA) constructed in λgt11. The library was plated using *Escherichia coli* strain Y1090 onto five 150-mm dishes at a density of 50,000 plaque-forming u/plate. The plates were incubated at 42°C for 3.5 h, and then the production of fusion proteins was induced by the addition of IPTG-containing nitrocellulose filters (Millipore Corp., Bedford, MA) and lowering the incubation temperature to 37°C. The plates were incubated for an additional 3.5 h and then were removed, washed in TBS (20 mM Tris HCl pH 7.5, 150 mM NaCl) and stored overnight at 4°C in 2% blocking reagent made up in TBS (Boehringer Mannheim Corp., Indianapolis, IN). A duplicate set of nitrocellulose filter lifts was taken and processed. The filters were reacted overnight at 4°C with the anti-NH$_2$-$\alpha1(V)$ antibodies diluted 1:400 in 2% blocking reagent. The plaques which had bound the antibodies were identified employing an alkaline phosphatase antibody detection kit (pico-Blue Immunoscreening Kit) used according to the supplier's (Stratagene Inc., La Jolla, CA) direction. Positive plaques were cloned and identified using the same protocol. In preparation for sequencing, inserts in the λgt11 vector were amplified by PCR using vector-specific primers and cloned into the plasmid vector pCR1000 using the TA cloning system (Invitrogen, San Diego, CA).

By sequencing and size determination on agarose gels, none of the immunoisolated clones appeared to encode the entire amino end of the $\alpha1(V)$ chain (see Results). Therefore, one of these clones (V-2.3.2) was used to screen a 13-d embryonic chicken corneal cDNA library we had previously constructed in the vector λZAPII. For this, the cDNA insert (1.6 kb) of clone V-2.3.2 was amplified by PCR using a GeneAmp gene-amplification kit (Perkin-Elmer/Cetus Corp., Norwalk, CT) and then electro-eluted from a 1% Agarose gel. The purified insert was labeled with ^{32}P by nick translation and used to probe plaque lifts of the λZAP cDNA library. Hybridizations were done overnight at 65°C in 5× SSC containing 1% N-lauroyl-sarcosine. Four washings were done at the same temperature: two in 3× SSC, 0.5% N-lauroylsarcosine followed by two in 3× SSC alone. Bound radioactivity was visualized by autoradiography. Positive plaques were cloned and the insert-containing Bluescript plasmids were rescued by helper-phage mediated in vivo excision.

Sequencing was performed using either the Sanger dideoxy method (42) as modified for the Sequenase kit (US Biochemical Corp., Cleveland, OH) or with Taq polymerase using a dsDNA Cycle Sequencing System kit (GIBCO BRL, Gaithersburg, MD). For the cycle sequencing method, the parameters used for the amplification cycles were those recommended by the supplier, except that the timing for the denaturation and elongation steps was increased from 30–50 s. The insert-containing plasmids to be sequenced by the dideoxy method were prepared by the lysozyme "Boiling Method"; those to be sequenced by the PCR method were prepared by the "Alkaline Lysis Method" (35).

Oligonucleotide primers were synthesized on an automated synthesizer (PCRMATE, model 391, Applied Biosystems, Inc., Foster City, CA). They were purified with oligonucleotide purification cartridges (Applied Biosystems, Inc.).

The sequenced samples were electrophoresed on 5% denaturing acrylamide wedge gels and visualized by autoradiography. The autoradiograms were read and the data recorded with the aid of the DNA Proscan Scanner and Sequencing Package (DNA Proscan, Inc., Nashville, TN). Some of the analyses of the data were performed using the GCG software package and Chou-Fasman prediction (6).

Results

Immunohistochemistry

Antibodies. The polyclonal antibodies employed (referred to as the anti-NH$_2$-$\alpha1(V)$ antibodies) are directed against epitopes present in the NH$_2$-terminal domain retained by the completely-processed (tissue) form of the $\alpha1(V)$ chain (see Materials and Methods). The two anti-type V mAbs used (V-DH2 and V-AB12) have been previously characterized as being directed against triple-helical, collagenous determinants within the type V molecule (34, 36).

Rotary Shadowing. For certain of the proposed studies, it was desirable to know whether the determinants to which these antihelical mAbs bind are located near the NH$_2$- or the COOH-terminal ends of the molecule. Thus, we performed rotary shadowing of their binding to native type V molecules with the attached NH$_2$-terminal domain. In rotary shadowed preparations, the NH$_2$-terminal domain retained with the native type V molecule appears to be a multidomain structure consisting of a kink, followed by a short rod, terminating in a ball (Fig. 1, *A* and *B*). It is ~17-nm long, a size which fits well with previous determinations made from rotary shadowed preparations (2) and from segment-long-spacing crystallites (4). The anti-type V mAbs (Fig. 1, *C* and *D*) bind near the opposite end of the molecule placing their binding sites ~15% from its COOH-terminal end.

Immunofluorescence. Immunofluorescence analyses were used to compare the corneal tissue reactivities of the NH$_2$-terminal anti-NH$_2$-$\alpha1(V)$ antibody with that of the antibodies to the helical sites. Two different ages of embryonic corneas were examined, 11-d embryonic corneas in which the corneal stroma is hydrated, swollen and rapidly undergoing matrix deposition, and 17-d embryonic corneas in which the stroma is mature and has attained its compact state.

At both times of development, the anti-NH$_2$-$\alpha1(V)$ antibodies and the anti-helical antibodies produced strikingly different patterns of immunoreactivity in untreated tissues. Whereas the anti-helical antibodies require a pretreatment resulting in fibril disruption, as previously described (34), the anti-NH$_2$-$\alpha1(V)$ antibodies reacted with the normal, untreated corneal stroma. This can be seen by comparing the fluorescence micrographs in Fig. 2, *A* and *B* (reacted with the different antibodies in the absence of tissue unmasking) with the micrographs in Fig. 2, *C* and *D* (reacted with the antibodies after unmasking by fibril disruption). Fig. 2, *E* and *F*, show control sections which were treated respectively with a mAb against type IV collagen and preimmune rabbit serum. The mature, 17-d corneal stroma behaves identically (data not shown).

Immunoelectron Microscopy. By immunoelectron microscopy of corneal stromas, we previously demonstrated that the antibodies against the helical epitopes on type V collagen reacted with the 24-nm striated fibrils, but again, only after fibril structure was partially disrupted (data not shown, but see 3, 13). The anti-NH$_2$-$\alpha1(V)$ antibodies also react with the 24-nm striated fibrils (Figs. 3, *B* and *D*) but, consistent with the immunofluorescence results, they need not be
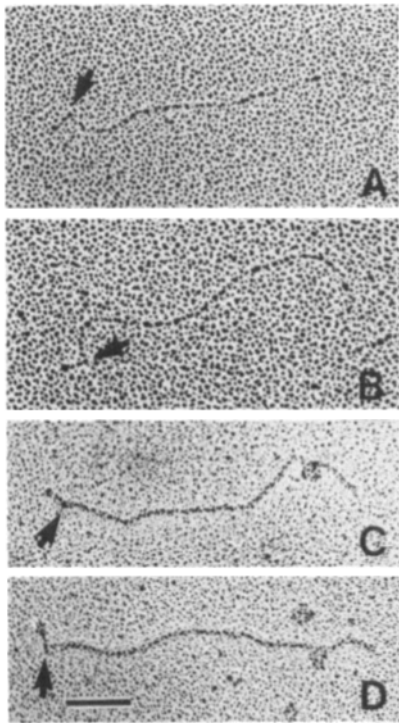
Figure 1. Electron micrographs of rotary shadowed preparations of monoclonal antibodies bound to native type V collagen molecules extracted from corneas of lathyritic chicken embryos. A and B show the type V molecules themselves. C and D show molecules that had been reacted with mAb directed against the triple-helical domain of the molecule. Arrows point to the kink demarcating the start of the NH₂ terminal domain. Note that the antibody binds near the opposite end of the molecule. Bar, 50 nm.

disrupted. This was observed whether pre-embedding im-munolocalization of conventional plastic sections (Fig. 3 B) or postembedding localization of ultrathin cryosections was employed (Fig. 3, C and D).

These data are consistent with the NH₂-terminal domain of the α1(V) chain being retained after the molecule has become incorporated into fibrils and, in addition, demonstrate that at least a portion of this domain is exposed on the fibril's surface.

Immunoisolation and Sequencing of cDNA Clones for the Complete NH₂ terminal Propeptide

To determine the structure of the NH₂-terminal domain, and that of the entire amino end of the α1(V) chain, we used the anti-NH₂-α1(V) antibody to immunoisolate cDNA clones, which were then sequenced. Initially, we screened two cDNA libraries, one constructed in λgt11 from 10-d chicken embryo mRNA, and one constructed in λZAPII from 13-d chicken embryo corneal mRNA. Ten positive plaques were identified and cloned from the λgt11 library; the λZAP library gave none. Several positive clones were partially sequenced. Each contained a collagenous sequence at its 3'-end and a noncollagenous sequence at its 5' end (Fig. 4). None, however, appeared to be large enough to encode a sequence of the size required for the entire amino end, as estimated from the biosynthetic studies on the molecule (8), and none contained a sequence encoding an in-frame initiation methionine codon.
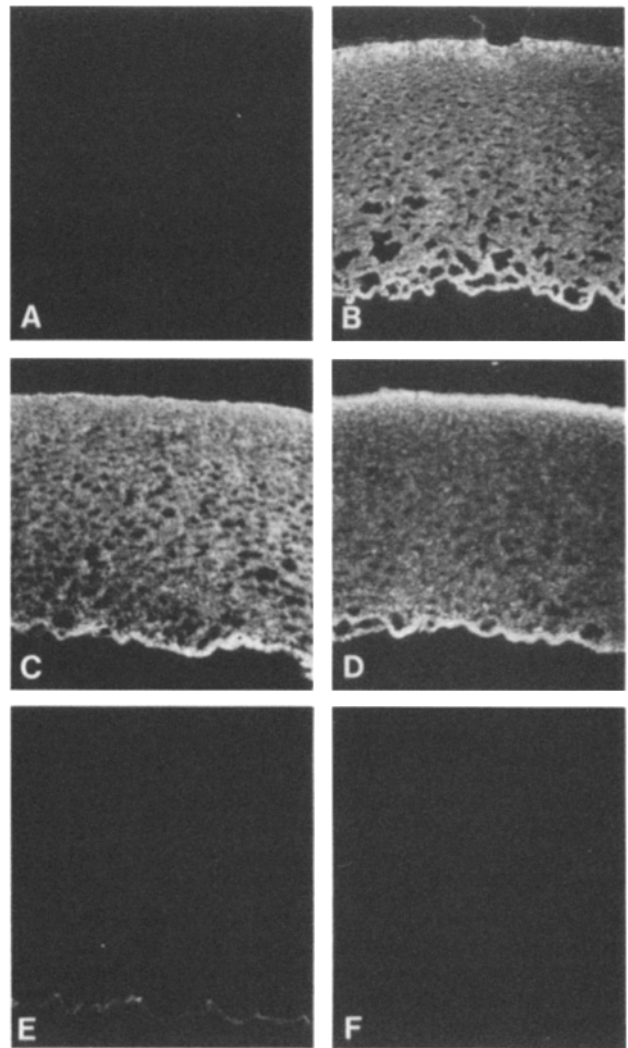

Figure 2. Immunofluorescence micrographs of corneal sections from 11-d chicken embryos reacted with anti-type V mAbs (V-AB12 and V-DH2) directed against the helical domain (A and C), or with anti-NH₂-α1(V) antibodies (B and D). C and D had been pretreated with 0.1 N HAc to disrupt fibrillar structure (34); A and B had no pretreatment. As controls, the section in E was reacted with an anti-type IV collagen antibody (14); the section in F was reacted with preimmune rabbit serum.

One of these clones, V-2.3.2., was used to rescreen the corneal λZAP library by hybridization. By this method, five positive plaques were now identified. Upon sequencing the largest of these clones, V105.1.1 (Fig. 4), a methionine codon was found, one base in from the 5' end. This clone was then sequenced past the point where it entered the major collagenous domain. Overlaps occurring between the λgt11 and the λZAP clone inserts were identical in sequence.

The sequence of clone V105.1.1 provided the conceptual amino acid sequence of the entire amino end of the chicken α1(V) chain. (The nucleotide sequence can be obtained from GenBank under the accession number Z21816.) In Fig. 5, the amino acid sequence is compared with that of the human (18) which has not become available.

Structure and Domains. As can be seen from the translated amino acid sequence, the entire amino end of the chicken α1(V) chain consists of 554 amino acid residues to the first triple-helical glycine of the main collagenous do-
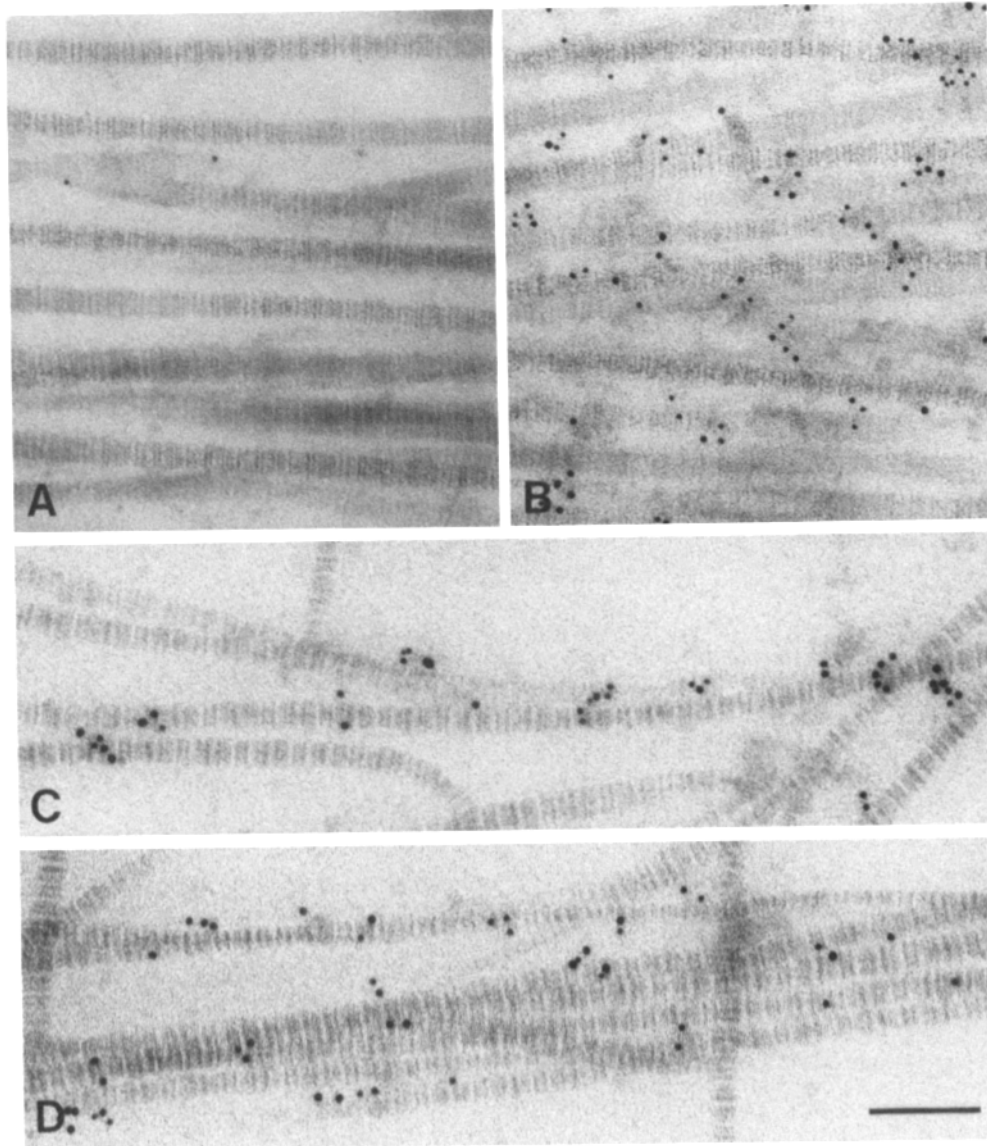
Figure 3. Immunoelectron micrographs of corneal sections reacted with either preimmune serum (*A*) or antisera directed against the NH₂-terminal domain of the α1(V) chain (*B–D*). All sections were reacted with a colloidal gold-labeled secondary antibody. *A* and *B* are conventional plastic sections of pre-embedded preparations. *C* and *D* are ultrathin cryosections employing postembedding localization. Bar, 100 nm.
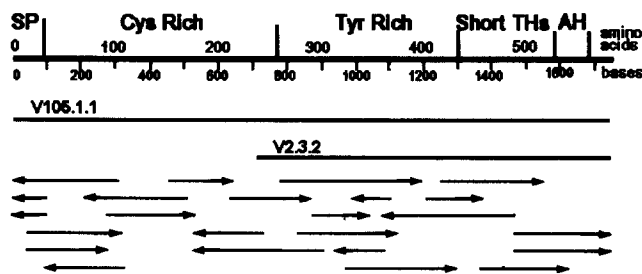

Figure 4. A schematic diagram showing the structure of the entire amino end of the pro-α1(V) chain, deduced from cDNA sequencing. Shown also is the strategy employed to obtain the sequence. V2.3.2 and V105.1.1 are two of the cDNA clones. V2.3.2 was obtained by immunoisolation; V105.1.1 was obtained by hybridization screening with V2.3.2. Clone V105.1.1 provided an uninterrupted sequence of the entire amino end of the pro-α1(V) chain. The major regions of the domain are designated: (SP) signal peptide, (*Cys Rich*) cysteine rich, (*Tyr Rich*) tyrosine rich, (*Short THs*) short triple helices, and (*AH*) α-helix. The arrows show the sequencing steps used to obtain the entire sequence of clone V105.1.1.

main. This is four residues more than in the human. From the translated sequence, we have deduced that this peptide is comprised of several regions, schematically diagrammed in Fig. 4.

Starting at the NH₂ terminus, the first of these is the signal peptide which consists of 33 amino acids, three less than the human. It contains a number of substitutions compared to the human, but at critical positions, the substituted amino acids maintain the characteristics required for a signal peptide (48).

The next major region is cysteine-containing, extends approximately to residue 249, and is highly conserved when compared to the human (88% identity). It contains all four cysteines of the NH₂-terminal propeptide as well as the single consensus sequence for glycosylation. All of these are conserved across the two species. By Chou-Fasman prediction (6), this cysteine-containing region has numerous intermixed hydrophobic and hydrophilic stretches, and numerous regions of beta sheet.

The high degree of interspecies amino acid conservation

```
  1 MDTHTRWKRRSWIR...NWQLHVALVLLGAAALGRAAEPADLLKVLDFHN  47
    ||  ||||| || |        ||| |   ||| ||||||||||||||
  1 MDVHTRWKARSALRPGAPLLPPLLLLLLWAPPPSRAAQPADLLKVLDFHN  50

 48 LPDGITRTTGFCTSRRSSKEADVAYRVTKDAQLSAPTKQLYPASPFPEDF  97
    ||||| |||||  ||||| |||||||||| |||||||||||||| |||||
 51 LPDGITKTTGFCATRRSSKGPDVAYRVTKDAHVSAPTKQLYPASAFPEDF 100

 98 SILTTVKAKKGGQSFLISIYNEQGIQQIGVEMGRSPVFLYEDHTGKPGPE 147
    |||||||||||| | || |||||||||||| | |||||||||||||||||
101 SILTTVKAKKGSQAFLVSIYNEQGIQQIGLELGRSPVFLYEDHTGKPGPE 150

148 DYPLFRGINLADGKWHRVAISVQKKNVTLILDCKKKITKFLDRSDHPIID 197
    ||||||||| |||||| | ||| ||||||| ||||||||||||||| ||
151 DYPLFRGINLSDGKWHRIALSVHKKNVTLILDCKKKTTKFLDRSDHPMID 200

198 VNGIIVFGTRILDEEVFEGDIQQLLIVADPRAAHDYCEHYSPDCDTAVPD 247
    |||||||||||||||||||||||||| | | ||| ||||||||||||||
201 INGIIVFGTRILDEEVFEGDIQQLLFVSDHRAAYDYCEHYSPDCDTAVPD 250

248 APQSQDPNQDEYYTDGEGEGDTYYYEYPYYEDVDEAVKPEAPTTKPAPPG 297
    |||||| |||||| | ||| |||||||||||||     | |   |
251 TPQSQDPNPDEYYTEGDGEGETYYYEYPYYEDPEDLGKEPTPSKKPVEAA 300

298 VAAGERPETKQDYPXPTPSPEAGNPSRQT.KGAAPVDDPLVDEYNYETIN 346
     | ||       || |   |  || |      ||
301 KETTEVPEE......LTPTPTEAAPMPETSEGAGKEEDVGIGDYDYVP.S 343

347 EEYFTPLPYEDINYNEEVDPQGGLTENAVEAELPTSTVITYNETD....A 392
    | | || || | |         ||  |||| | |
344 EDYYTPSPYDDLTYGEGEENPDQPTDPGAGAEIPTSTADTSNSSNPRPPP 393

393 AQGGDDLDKDFTEETIKEYDGNYY...YYDRTVSP.DIGPGMPANQDTIYE 439
     | |||  ||||||| |||  |  | |||  || ||  ||||||||||||
394 GEGADDLEGEFTEETIRNLDENYYDPYYDPTSSPSEIGPGMPANQDTIYE 443

440 GIGGPRGEKGQKGEPAIIEPGMLVEGPPGPEGPAGLPGPPGPTGPVGLMG 489
    ||||||||||||||||||||||||:||||||||||||||||||  ||  |
444 GIGGPRGEKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTMGPTGQVG 493

490 DPGERGPPGRPGLPGADGLPGPPGTMLMLPFRFSGGGDAGSKGPMVSAQE 539
    |||||||||||||||||||||||||||||||||  |||||||||||||||
494 DPGERGPPGRPGLPGADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQE 543

540 AQAQAILQQARLALRGPAGPMGLTGRPGPIGPQGSGG             576
    |||||||||||||||||||||||||||| |  ||||||
544 SQAQAILQQARLALRGPAGPMGLTGRPGPVGPPGSGG             580
```

Figure 5. Comparison of the translated amino acid sequences for entire amino end of the pro-α1(V) chain of the chicken (top sequence) and the human (18) (lower sequence). Arrowheads demarcate the end of the signal peptide and the start of the major triple helix. Cysteines are designated with asterisks; alanine-glutamine/proline-glutamine sequences have overhead lines, and the glycosylation site is boxed. Tryosines have dots above them; those which are likely to be sulfated have circles surrounding the dots.

Table I. Propeptide Cleavage Sites (PQ and AQ) and Amino Acids at Positions −2 and −3

α1(I)	Bovine	F A P Q
α2(I)	Chick	F A A Q
α1(II)	Xenopus	F A A Q
α2(V) (proposed)	Human	F S A Q
α1(III)	Bovine	Y A P Q
α1(V) (proposed)	Chick (249–250)	D A P Q

The carboxyl-most regions, which again are highly conserved between the species (94% identity), consist of short collagenous stretches (three) interrupted by noncollagenous ones (see Fig. 4 "Short THs"). The last of the noncollagenous regions is short and has a strong tendency towards α-helix formation (Fig. 4, AH). It immediately precedes the major collagenous domain of the molecule, and contains a potential propeptidase cleavage site, the alanine-glutamine at position 537–538, but one which we do not think is utilized (addressed next, and see Discussion).

Potential Propeptidase Cleavage Sites

To obtain information on which portion of the amino end of the α1(V) chain is removed as the NH_2-propeptide and which portion is retained with the completely processed molecule as the NH_2-terminal domain, we examined the sequence for alanine-glutamine and proline-glutamine sequences. These are the sites in the chains of procollagen types I-III that are thought to be cleaved by NH_2-terminal procollagen peptidases (23, 40, 44, 50). We also examined the amino acids at positions −2 and −3 from these sites, since Morikawa et al. (39) have provided evidence that a phenylalanine at position −3 appears to be crucial for the action of the NH_2-terminal type I procollagen propeptidase. Seven alanine-glutamine or proline-glutamine sequences were found (demarcated in Fig. 5 by overhead lines), but only three of these had counterparts in the human sequence (one at position 249–250, a second at position 537–538 and a third at 542–543). No site, however, has an associated phenylalanine at position −3 (Table I). Comparing these data with the protein processing data of others suggests to us that the site at position 249–250 is likely to be the one utilized during the processing of type V collagen (see Discussion).

rapidly decreases in the next region (46% identity). This tyrosine-rich region begins at a putative propeptidase cleavage site, the proline-glutamine at positions 249–250 (see below and Discussion). It extends for ~175 residues to a site ~15 amino acid residues before a region consisting of three short triple helices. When compared to the human sequence, this region is found to contain numerous amino acid substitutions, one insertion of six amino acids not found in the human and three deletions of one, two, and four amino acids respectively. Of the tyrosines in this region, at least four, and possibly six, are likely sites for sulfation, based on the consensus features of sulfation sites (24, 25). All except one are conserved between the two species. This region is also highly hydrophilic as determined by Chou-Fasman prediction and is likely to be retained with the completely processed type V collagen molecule (see Discussion).

Discussion

Fibril Structure

The morphological data from both immunofluorescence and immunoelectron microscopy demonstrate that in the heterotypic type I + V corneal collagen fibrils, a portion of the amino end of the pro-α1(V) chain is retained with the type V molecule as the NH_2-terminal domain. The data also demonstrate that at least some epitopes in this domain are exposed on the surface of such fibrils. The extent to which this domain and its adjacent collagenous domain are exposed is not yet known. We do know, however, that a portion of the helical domain of the type V collagen molecules is buried within these fibrils. In the least, this includes the region to which our anti-helical mAbs bind. As shown in the present study, this is near the COOH-terminal end of the molecule.
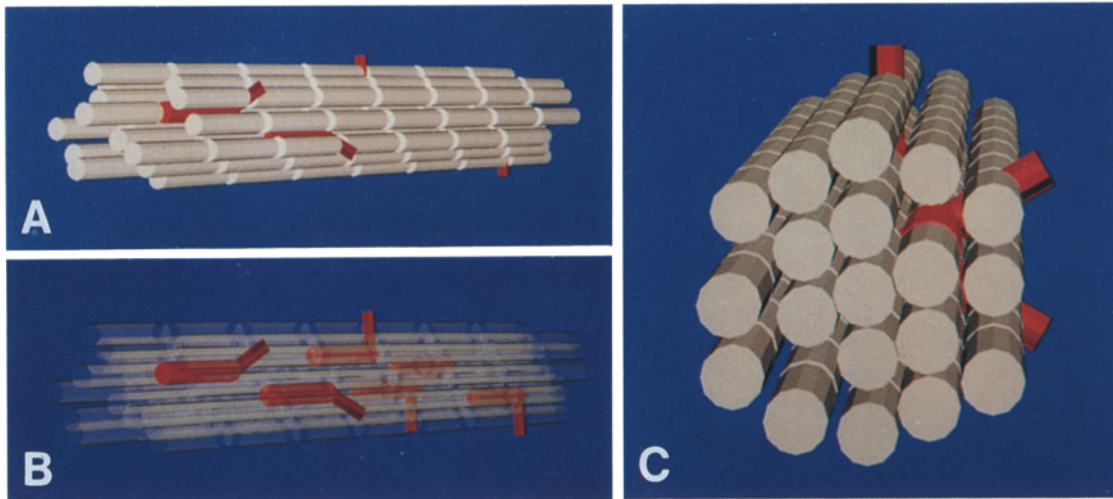
Figure 6. Three views of a proposed model for the arrangement of type V collagen molecules within a heterotypic type I + V collagen fibril. *A* and *B* show the fibril in longitudinal view; *C* shows the fibril in cross-section. The type I molecules are colored beige; the type V molecules are colored red. The type V molecules have the retained NH_2-terminal domain extending perpendicular to the long axis. The type V molecules are arranged such that their triple-helical domain lies within the interior of the fibril and their NH_2-terminal domain extends through a hole zone to the fibril surface. In *B*, the type I molecules have been rendered transparent, thus, allowing full visualization of the interior type V molecules.

The buried region also includes site(s) which are cleaved by a type V collagen-degrading collagenase (13). Both the anti-helical mAb-binding sites and the type V collagen-degrading collagenase cleavage site(s) can be exposed by the removal of some of the type I collagen with a collagenase which selectively cleaves type I molecules (13). This suggests that within type I + V heterotypic collagen fibrils at least a portion of the triple-helical domain of each type V collagen molecule, and possibly the entire triple-helical domain of each, lies interior to a surface of type I collagen molecules.

At least two models can be proposed for the structure of a heterotypic type I + V collagen fibril which are consistent with these data. In one model, both types of collagen molecules would be arranged overlapping and at an angle to the fibrillar axis, as has been previously suggested for collagen fibrils in general (5). If the NH_2-terminal domain of the type V molecules were directed outward, it would explain the presence of this region at the fibril surface. This model, however, also predicts that one end of the type I collagen molecules would be buried within such a fibril, and this is not consistent with our observations. Two anti-type I collagen antibodies, directed against epitopes near opposite ends of the molecule, both label fibrils at intervals along the surface indicating that both ends of this molecule are exposed on the fibril surface (unpublished observations).

An alternative model which we consider more likely is diagrammed in Fig. 6. In this model, the type I and type V molecules are arranged parallel to one another. This requires that the entire triple-helical domain of the type V molecule be buried within the fibril, and that type I molecules are completely exposed along the fibril's surface. In such an arrangement, for the retained portion of the NH_2-terminal domain of the type V collagen molecules to be exposed at the fibril's surface, they must extend outward through the hole zones of the fibril. Hole zones are an inherent structural element of striated collagen fibrils and are formed by the incomplete overlapping of collagen molecules when arranged in a quarter-stagger, fibrillar array (19). The rotary shadow-

ing evidence and sequencing data (discussed below) suggest that the NH_2-terminal domain of the $\alpha 1(V)$ chain consists of a flexible hingelike region followed by a short collagenous sequence ending in a tyrosine-rich region. The hinge could allow the peptide to project away from the major axis of the molecule and the short triple-helical domain could serve as an extension through the hole zone, placing the tyrosine-rich domain at the fibril's surface.

Regardless of which model, if either, is correct for the heterotypic type I + V fibrils, the data are consistent with the proposition that a region of the NH_2-terminal domain on the fibril surface regulates fibril diameter by sterically preventing the further addition of collagen molecules. It is also possible that the presence of highly charged sulfate groups on the tyrosines in this region (9, 10) also inhibits further accretion of collagen molecules to the surface.

Sequencing of cDNAs

The conceptual translation product derived from the cDNA clones encoding the amino end of $\alpha 1(V)$ chain has provided information about the size of the retained NH_2-terminal domain. It also explains why this occurs.

The sequencing data, when coupled with analyses of the protein structure and processing of type V collagen (4, 8, 11, 30) suggest that the portion of amino end which is retained with molecules incorporated into a fibril is likely to derive from cleavage of the proline-glutamine site at position 249–250. The retained NH_2-terminal domain would therefore include the tyrosine-rich and short triple-helical regions.

Fessler et al. (9, 10) have reported that tyrosines in the amino end of the $\alpha 1(V)$ chain are sulfated, and that most of these are retained after the molecule has undergone complete processing. The sequencing data show that all of the tyrosines capable of undergoing sulfation (24, 25) are contained within the tyrosine-rich region (see also reference 45) adjacent to this putative propeptide cleavage site. Even

though the amino acid residues in this region are not highly conserved between the chick and the human (46% identity), all but one of the potential tyrosine sulfation sites are conserved. Also, its highly hydrophilic nature would be consistent with a location at the surface of a fibril.

Analyses of the size of the NH_2-terminal domain retained with the completely processed $\alpha1(V)$ chain have produced disparate results. The basis for these differences is unknown, but one possibility is that developmental and tissue-specific differences exist in posttranslational modifications, such as in the sulfation of the tyrosines. Broeck et al. (4) examined the type V collagen extracted from bone and other tissues of mature chickens that had been made lathyritic over a period of weeks. All of the type V molecules contained $\alpha1$ chains with a bacterial collagenase resistant peptide of 18 KD as determined by SDS PAGE.

Germane to the present study, we have extracted a similar size chain from lathyritic chicken corneas (unpublished data). The 18-KD peptide fits well with the size of the collagenase resistant region ($\sim$190 amino acid residues) one would expect if the putative peptidase cleavage site utilized during procollagen processing were the one at position 249–250. If so, then the short collagenous regions are also retained with the processed molecule. By rotary shadowing analyses, the type V molecules extracted from lathyritic corneas show a substructure within the NH_2-terminal domain consisting of a hinge, followed by a short rodlike stretch and finally a ball. The conceptual amino acid sequence suggests that the hinge is located within a short (20 amino acid) stretch adjacent to the major triple helical domain of the molecule. By Chou-Fasman prediction, this stretch has a strong tendency towards α-helix formation. The short, rodlike stretch is likely formed by the three small collagenous regions with the ball being the tyrosine-rich domain.

If the proline-glutamine at position 249–250 is the cleavage site which is utilized, it most likely requires a different NH_2-procollagen peptidase than the one utilized for type I procollagen. The specific sequence required for cleavage by the propeptidase which cleaves the $\alpha1(I)$ and the $\alpha2(I)$ chains of type I collagen is the only one that has been examined experimentally (39). The results suggest that in addition to the amino acids immediately flanking the cleavage site, the phenylalanine located at position −3 is also required. Presumably, in type III collagen, the substitution of a tyrosine at −3 in the $\alpha1(III)$ chain is the reason that this molecule is cleaved not by the type I collagen propeptidase but by another enzyme (20, 40). Theoretically, other amino acids surrounding the peptide bond which is cleaved might also be important in determining specificity, as has recently been shown for the vertebrate collagenase cleavage site in type I collagen (51). It seems likely that yet another procollagen peptidase is involved in the processing of the NH_2-terminal end of the $\alpha1(V)$ chain.

Comparison of the sequences for the pro-$\alpha1(V)$ and pro-$\alpha2(V)$ chains of the type V procollagen molecule show that the amino end of the $\alpha2$ chain is less than half the size of that of the $\alpha1$ chain. Moreover, in the amino end of the pro-$\alpha2(V)$ chain there is no site homologous to that proposed to undergo cleavage in the pro-$\alpha1(V)$. In the pro-$\alpha2(V)$ chain, however, a complete consensus sequence does exist for cleavage by the type I collagen propeptidase (49), but this site is located near the major triple helix (i.e., in a telopep-

tidelike region). Studies on processing of the type V procollagen molecule suggest that cleavage of the $\alpha2(V)$ propeptide is more rapid than is that of the $\alpha1(V)$ chain. Within a protype V molecule of chain composition [pro-$\alpha1(V)$]$_2$pro-$\alpha2(V)$, cleavage of the pro-$\alpha2$ chain at the NH_2 telopeptidelike site, but not the two pro-$\alpha1$ chains, might be a factor in producing the flexibility for forming the kink observed to exist between the NH_2-terminal domain and the major triplehelical domains of type V molecules viewed by rotary shadowing.

Whether the putative cleavage site in the amino end of pro-$\alpha1(V)$ chain requires a unique procollagen peptidase remains to be determined, but there is a precedent for unique enzymes acting on type V collagen. Another enzyme-mediated alteration of type V collagen, its degradation, employs a collagenase different from the one that cleaves most of the other fibrillar collagens (13, 21). The isolation of a procollagen peptidase specific for the putative cleavage site in the $\alpha1(V)$ chain would provide a critical test for some of the suggestions raised here.

References

1. Bachinger, H. P., K. J. Doege, J. P. Petschek, L. I. Fessler, and J. H. Fessler. 1982. Structural implications from an electronmicroscopic comparison of procollagen V with procollagen I, pC-collagen I, procollagen IV, and a Drosophila procollagen. J. Biol. Chem. 257:14590–14592.
2. Birk, D. E., J. M. Fitch, J. P. Babiarz, K. J. Doane, and T. F. Linsenmayer. 1990. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. J. Cell Sci. 95:649–657.
3. Birk, D. E., J. M. Fitch, J. P. Babiarz, and T. F. Linsenmayer. 1988. Collagen type I and type V are present in the same fibril in the avian corneal stroma. J. Cell Biol. 106:999–1008.
4. Broek, D. L., J. Madri, E. F. Eikenberry, and B. Brodsky. 1985. Characterization of the tissue form of type V collagen from chick bone. J. Biol. Chem. 260:555–562.
5. Chapman, J. A. 1989. The regulation of size and form in the assembly of collagen fibrils in vivo. Biopolymers. 28:1367–1382.
6. Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47:45–148.
7. Doane, K. J., J. P. Babiarz, J. M. Fitch, T. F. Linsenmayer, and D. E. Birk. 1992. Collagen fibril assembly by corneal fibroblasts in three-dimensional collagen gel cultures: Small-diameter heterotypic fibrils are deposited in the absence of keratan sulfate proteoglycan. Exp. Cell Res. 202:113–124.
8. Fessler, J. H., and L. I. Fessler. 1987. Type V collagen. In Structure and Function of Collagen Types. R. Mayne and R. E. Burgeson, editors. Academic Press, Orlando. 81–97.
9. Fessler, L. I., S. Brosh, S. Chapin, and J. H. Fessler. 1986. Tyrosine sulfation in precursors of collagen V. J. Biol. Chem. 261:5034–5040.
10. Fessler, L. I., S. Chapin, S. Brosh, and J. H. Fessler. 1986. Intracellular transport and tyrosine sulfation of procollagens V. Eur. J. Biochem. 158:511–518.
11. Fessler, L. I., C. A. Kumamoto, M. E. Meis, and J. H. Fessler. 1981. Assembly and processing of procollagen V (AB) in chick blood vessels and other tissues. J. Biol. Chem. 256:9640–9645.
12. Fessler, L. I., N. Shigaki, and J. H. Fessler. 1985. Isolation of a new procollagen V chain from chick embryo tendon. J. Biol. Chem. 260:13286–13293.
13. Fitch, J. M., D. E. Birk, A. Mentzer, K. A. Hasty, C. Mainardi, and T. F. Linsenmayer. 1988. Corneal collagen fibrils: dissection with specific collagenases and monoclonal antibodies. Invest. Ophthalmol. Vis. Sci. 29:1125–1136.
14. Fitch, J. M., E. Gibney, R. D. Sanderson, R. Mayne, and T. F. Linsen-

mayer. 1982. Domain and basement membrane specificity of a monoclonal antibody against chicken type IV collagen. *J. Cell Biol.* 95:641–647.

15. Fitch, J. M., J. Gross, R. Mayne, B. Johnson Wint, and T. F. Linsenmayer. 1984. Organization of collagen types I and V in the embryonic chicken cornea: monoclonal antibody studies. *Proc. Natl. Acad. Sci. USA.* 81:2791–2795.

16. Fleischmajer, R., B. R. Olsen, R. Timpl, J. S. Perlish, and O. Lovelace. 1983. Collagen fibril formation during embryogenesis. *Proc. Natl. Acad. Sci. USA.* 80:3354–3358.

17. Fleischmajer, R., J. S. Perlish, and R. Timpl. 1985. Collagen fibrillogenesis in human skin. *Ann. NY Acad. Sci.* 460:246–257.

18. Greenspan, D. S., W. Cheng, and G. G. Hoffman. 1991. The pro-α1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-α1(XI) collagen chain. *J. Biol. Chem.* 266:24727–24733.

19. Gross, J. 1974. Collagen biology: structure, degradation, and disease. *Harvey Lect.* 6B:351–432.

20. Halila, R., and L. Peltonen. 1986. Purification of human procollagen type III N-proteinase from placenta and preparation of antiserum. *Biochem. J.* 239:47–52.

21. Hasty, K. A., T. F. Pourmotabbed, G. I. Goldberg, J. P. Thompson, D. G. Spinella, R. M. Stevens, and C. I. Mainardi. 1990. Human neutrophil collagenase. A distinct gene product with homology to other matrix metalloproteinases. *J. Biol. Chem.* 265:11421–11424.

22. Hodge, A. J., and J. A. Petruska. 1963. Recent studies with the electron microscope on ordered aggregates of the tropocollagen molecule. *In* Aspects of Protein Structure. G. N. Ramachandran, editor. Academic Press, New York. 289–301.

23. Horlein, D., P. P. Fietzek, E. Wachter, C. M. Lapiere, and K. Kuhn. 1979. Amino acid sequence of the aminoterminal segment of dermatosparactic calf skin procollagen type I. *Eur. J. Biochem.* 99:31–38.

24. Hortin, G., J. Folz, J. I. Gordon, and A. W. Strauss. 1986. Characterization of sites of tyrosine sulfation in proteins and criteria for predicting their occurrence. *Biochem. Biophys. Res. Commun.* 141:326–333.

25. Huttner, W. B. 1987. Protein tyrosine sylfation. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 12:361–363.

26. Ibrahim, J., and J. J. Harding. 1989. Pinpointing the sites of hydroxylysine glycosides in peptide α1-CB7 of bovine corneal collagen, and their possible role in determining fibril diameter and thus transparency. *Biochim. Biophys. Acta.* 992:9–22.

27. Kadler, K. E., D. J. S. Hulmes, Y. Hojima, and D. J. Prockop. 1990. Assembly of type I collagen fibrils *de novo* by the specific enzymic cleavage of pC collagen: the fibrils formed at about 37°C are similar in diameter, roundness, and apparent flexibility to the collagen fibrils seen in connective tissue. *Ann. NY Acad. Sci.* 580:214–224.

28. Keene, D. R., L. Y. Sakai, H. P. Bachinger, and R. E. Burgeson. 1987. Type III collagen can be present on banded collagen fibrils regardless of fibril diameter. *J. Cell Biol.* 105:2393–2402.

29. Keene, D. R., L. Y. Sakai, R. E. Burgeson, and H. P. Bachinger. 1987. Direct visualization of IgM antibodies bound to tissue antigens using a monoclonal anti-type III collagen IgM as a model system. *J. Histochem. Cytochem.* 35:311–318.

30. Kumamoto, C. A., and J. H. Fessler. 1981. Propeptides of procollagen V (A,B) in chick embryo crop. *J. Biol. Chem.* 256:7053–7058.

31. Linsenmayer, T. F. 1991. Collagen. *In* Cell Biology of Extracellular Matrix. E. D. Hay, editor. Plenum Publishing Corp., New York. 7–44.

32. Linsenmayer, T. F., J. M. Fitch, J. Gross, and R. Mayne. 1985. Are collagen fibrils in the developing avian cornea composed of two different collagen types? Evidence from monoclonal antibody studies. *In* Biochemistry, Chemistry and Pathology of Collagen. R. Fleischmajer, B. R. Olsen, and K. Kuhn, editors. New York Academy of Sciences, New York. 232–245.

33. Linsenmayer, T. F., J. M. Fitch, and R. Mayne. 1984. Extracellular matrices in the developing avian eye: type V collagen in corneal and noncorneal tissues. *Invest. Ophthalmol. Vis. Sci.* 25:41–47.

34. Linsenmayer, T. F., J. M. Fitch, T. M. Schmid, N. B. Zak, E. Gibney, R. D. Sanderson, and R. Mayne. 1983. Monoclonal antibodies against chicken type V collagen: production, specificity, and use for immunocytochemical localization in embryonic cornea and other organs. *J. Cell Biol.* 96:124–132.

35. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1986. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 1.23–1.30.

36. Mayne, R., H. Wiedemann, M. H. Irwin, R. D. Sanderson, J. M. Fitch, T. F. Linsenmayer, and K. Kuhn. 1984. Monoclonal antibodies against chicken type IV and V collagens: electron microscopic mapping of the epitopes after rotary shadowing. *J. Cell Biol.* 98:1637–1644.

37. McLaughlin, J. S., T. F. Linsenmayer, and D. E. Birk. 1989. Type V collagen synthesis and deposition by chicken embryo corneal fibroblasts *in vitro. J. Cell Sci.* 94:371–379.

38. Mendler, M., S. G. Eich-Bender, L. Vaughan, K. H. Winterhalter, and P. Bruckner. 1989. Cartilage contains mixed fibrils of collagen types II, IX, and XI. *J. Cell Biol.* 108:191–198.

39. Morikawa, T., L. Tuderman, and D. J. Prockop. 1980. Inhibitors of procollagen N-protease. Synthetic peptides with sequences similar to the cleavage site in the pro alpha 1(I) chain. *Biochemistry.* 19:2646–2650.

40. Peltonen, L., R. Halila, and L. Ryhanen. 1985. Enzymes converting procollagens to collagens. *J. Cell. Biochem.* 28:15–21.

41. Romanic, A. M., E. Adachi, K. E. Kadler, Y. Hojima, and D. J. Prockop. 1991. Copolymerization of pNcollagen III and collagen I. pNcollagen III decreases the rate of incorporation of collagen I into fibrils, the amount of collagen I incorporated, and the diameter of the fibrils formed. *J. Biol. Chem.* 266:12703–12709.

42. Sanger, F., and A. R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94:441–448.

43. Scott, J. E. 1988. Proteoglycan-fibrillar collagen interactions. *Biochem. J.* 252:313–323.

44. Su, M.-W., H. R. Suzuki, J. J. Bieker, M. Solursh, and F. Ramirez. 1991. Expression of two nonallelic type II procollagen genes during *Xenopus laevis* embryogenesis is characterized by stage-specific production of alternatively spliced transcripts. *J. Cell Biol.* 115:565–575.

45. Takahara, K., Y. Sato, K. Okazawa, N. Okamoto, A. Noda, Y. Yaoi, and I. Kato. 1991. Complete primary structure of human collagen α1(V) chain. *J. Biol. Chem.* 266:13124–13129.

46. Tseng, S. C. G., D. Smuckler, and R. Stern. 1982. Comparison of collagen types in adult and fetal bovine corneas. *J. Biol. Chem.* 257:2627–2633.

47. Vaughan, L., M. Mendler, S. Huber, P. Bruckner, K. H. Winterhalter, M. I. Irwin, and R. Mayne. 1988. D-periodic distribution of collagen type IX along cartilage fibrils. *J. Cell Biol.* 106:991–997.

48. von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133:17–21.

49. Woodbury, D., V. Benson-Chanda, and F. Ramirez. 1989. Aminoterminal propeptide of human pro-alpha 2(V) collagen conforms to the structural criteria of a fibrillar procollagen molecule. *J. Biol. Chem.* 264:2735–2738.

50. Wozney, J., D. Hanahan, V. Tate, H. Boedtker, and P. Doty. 1981. Structure of the pro alpha 2 (I) collagen gene. *Nature (Lond.).* 294:129–135.

51. Wu, H., M. H. Byrne, A. Stacey, M. B. Goldring, J. R. Birkhead, R. Jaenisch, and S. M. Krane. 1990. Generation of collagenase-resistant collagen by site-directed mutagenesis of murine proα1(I) collagen gene. *Proc. Natl. Acad. Sci. USA.* 87:5888–5892.