Alternative Splicing of the Human α2(VI) Collagen Gene Generates Multiple mRNA Transcripts Which Predict Three Protein Variants with Distinct Carboxyl Termini*

Biagio Saitta†, David G. Stokes, Henrik Vissing‡, Rupert Timpl§, and Mon-Li Chun¶

From the Departments of Biochemistry and Molecular Biology and Dermatology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Jefferson Medical College, Philadelphia, Pennsylvania 19107, the §Brookdale Center for Molecular Biology, Mount Sinai Medical Center, New York, New York 10029, and the ¶Max-Planck-Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany

We recently reported the isolation and sequencing of two classes of human α2(VI) collagen cDNA clones which share common sequences for the first two-thirds of the molecule but contain a different sequence of either 607 or 887 base pairs at their 3′ ends (Chu, M.-L., Pan, T.-C., Conway, D., Kuo, H.-J., Glanville, R. W., Timpl, R., Mann, K., and Deutzmann, R. (1989) EMBO J. 8, 1939–1946). In the present study, we report the sequence of another cDNA clone, which is identical to one class of the previously isolated cDNAs except for a 283-base pair insertion between the common and variable regions. Together, the different classes of cDNAs, referred to as the α2C2, α2C2a, and α2C2a’ predict three variant α2 chains of type VI collagen with carboxyl globular domains of 429, 328, and 238 amino acid residues, respectively. In order to explore the mechanisms by which the variations are generated, we isolated and characterized the 3′ end of the human α2(VI) collagen gene. The carboxyl globular domain was found to be encoded by six exons which appear to delineate its structural subdomains. The exon/intron arrangement clearly demonstrated that the cDNA variants arose from alternative splicing events by mutually exclusive utilization of the last two exons in conjunction with the selective usage of an internal splice acceptor site in the penultimate exon. The presence of the corresponding mature mRNA transcripts (3.2–3.5 kilobase pairs (kb)) in human fibroblasts was shown by Northern blot hybridization, S1 nuclease protection assay, and the polymerase chain reaction. The results indicated that the α2C2 mRNA is the major species, whereas the α2C2a and α2C2a’ are the minor forms. Northern blot hybridization also revealed an α2(VI) collagen mRNA of 6.0 kb. This mRNA retained a 2.3-kb intron located between the two alternatively spliced exons and predicted a translational product that is the same as the α2C2a variant.

Type VI collagen is one of the microfibrillar components found in a wide variety of tissues. Several lines of evidence suggest that it acts as an adhesive protein and may play a role in anchoring cells, large collagen fibers, and other tissue components such as blood vessels and nerves, in the surrounding extracellular matrices (1–3). Type VI collagen molecules contain three genetically distinct polypeptides, the α1(VI), α2(VI), and α3(VI) chains, with molecular masses of about 140, 140, and 300 kDa, respectively (4–6). Recently, we have elucidated the complete primary structure of the human α2(VI) chain by determining the cloned cDNA sequence in conjunction with peptide sequencing of the protein (7, 8). The predicted amino acid sequence demonstrated three major structural domains: a central triple helical domain (395 amino acid residues) flanked by amino globular (234 residues) and carboxyl globular (429 residues) domains. Further examination of the sequences revealed that the two globular domains are composed of three homologous subdomains of about 200 amino acids each. One of the repeating units was found in the amino globular domain (N1) and the other two in the carboxyl globular domain (C1 and C2). The repeats share sequence homology with similar domains found in von Willebrand factor, a cartilage matrix protein, complement factors, and integrins (9–12). The same structural features have also been identified in the chick α2(VI) collagen chain (13). We have, in addition, isolated and characterized a different kind of human α2(VI) cDNA which encodes a variant α2(VI) chain with a distinct carboxyl terminus (8). This variant has a unique 98-residue segment in replacement of the second repeat in the carboxyl globular domain.

We have now sequenced another cDNA clone which represents a third variant encoding an even shorter α2(VI) chain that terminated shortly after the first repeat of the carboxyl globular domain. Two observations support the suggestion that these variant cDNAs reflect alternative splicing of a primary transcript. First, the three classes of cDNAs share a common region that comprises two-thirds of the α2(VI) collagen mRNA from the 5′ end, and the sequence divergence occurs at the same point in all three variants. Second, Southern blot analysis of human genomic DNA indicated that the α2(VI) chain is encoded by a single copy gene located on human chromosome 21 (14). To further address this question, we isolated and characterized the 3′ end of the human α2(VI) collagen gene. Here we present evidence indicating that the cDNA variants arise from alternative splicing events and that multiple α2(VI) collagen mRNAs are indeed found in human fibroblasts.

* This work was supported by the National Institutes of Health Grants AR-38912, AR-38923, AR-38188, and AR-39740 and the Deutsche Forschungsgemeinschaft Grant SFB 266. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Jefferson Medical College, 1020 Locust St., Philadelphia, PA 19107; Tel.: 215–955–4834; Fax: 215–955–5393.
‡ On leave of absence from Istituto di Biologia dello Sviluppo, CNR, Palermo, Italy.

44x764
Isolation and Characterization of cDNA and Genomic Clones—An α2(VI) cDNA clone, F221, was isolated from a human fibroblast cDNA library as described previously (7). A human cosmid library was constructed using leukocyte DNA and the pHV4 vector. This vector is a variant of the cosmid vector pLTC (15). The cloning site is flanked by the T3 and T7 RNA polymerase promoters. Total genomic DNA was partially digested with Sau3A and DNA fragments of 30–50 kb were isolated and ligated with the pHV4 vector as described (15). The ligated DNA was packaged with packaging extract obtained from Stratagene (La Jolla, CA), and the packaged cosmids were used to transfect Escherichia coli RecA strain EDB176 (16).

Approximately 4 x 10⁶ colonies, representing five human genomes, were plated and screened with a previously characterized α2(VI) cDNA library as described previously (7). A human cosmid library was constructed using leukocyte DNA and the pHV4 vector. This vector is a variant of the cosmid vector pLTC (15). The cloning site is flanked by the T3 and T7 RNA polymerase promoters. Total genomic DNA was partially digested with Sau3A and DNA fragments of 30–50 kb were isolated and ligated with the pHV4 vector as described (15). The ligated DNA was packaged with packaging extract obtained from Stratagene (La Jolla, CA), and the packaged cosmids were used to transfect Escherichia coli RecA strain EDB176 (16). The clone was further characterized by Southern blot analysis using subfragments or synthetic oligonucleotides derived from the cDNA. Appropriate genomic DNA fragments were subcloned into either the BlueScript vectors (Stratagene, La Jolla, CA) or the M13 vectors. The appropriate genomic DNA fragments were subcloned into either the BlueScript vectors (Stratagene, La Jolla, CA) or the M13 vectors. The DNA sequences were determined by the dideoxy chain termination method (16) with either single- or double-stranded DNA templates using a kit from United States Biochemical (Cleveland, OH) and α-35S-thio-dATP (Du Pont-New England Nuclear).

Northern Blot Analysis—RNA was isolated from established human fibroblasts (GM 3349) obtained from Coriell Institute for Medical Research, Camden, NJ. Total RNA was prepared in guanidinium isothiocyanate, extracted with phenol-chloroform (17) and poly(A)+ RNA in 80% formamide, 300 mM NaCl, 30 mM sodium acetate, pH 4.5 and treated for 15 min at 37 °C with 4 units of S1 nuclease (Promega, Madison, WI) and analyzed on a 6% acrylamide-urea gel.

RESULTS

Three Classes of the α2(VI) Collagen cDNA Clones—Two α2(VI) collagen cDNA variants have been isolated previously and characterized from human placenta and fibroblast cDNA libraries (8). The nucleotide sequence of these two variants were identical from their most 5' ends up to 2458 nucleotides downstream from the translation start site, whereas their 3' terminal 607- and 887-bp sequences were completely different. The abbreviations used are: kb, kilobase pair(s); bp, base pair(s); Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PCR, polymerase chain reaction; CS, connecting segments.

The predicted protein domains show the signal sequence (S), the repeating units in the amino globular domain (N1), and carboxyl globular domain (C1, C2), triple-helix (TH), short cysteine-rich connecting segments (CS1, CS2, CS3), and the alternative forms of the carboxyl terminus (C2α and C2α*). Indicated restriction enzyme sites are PstI (P), BamHI (B), XhoI (X), AccI (A), B, nucleotide and predicted amino acid sequence of the variable region of the α2C2α variant cDNA clone, F221. The variable region starts at nucleotide 2480, following the numbering described previously (8). Between the two triangles is the 293-bp segment not found in the α2C2α variant.
(Ref. 8 and Fig. 1A). Therefore, they share a common region containing coding sequences for the amino globular, triple helical, and the C1 subdomain of the carboxyl globular region. The variable regions encoded either a 199-residue C2 subdomain plus a 291-bp 3'-untranslated region or a 98-residue C2a subdomain plus a 514-bp 3'-untranslated region, respectively (8). Based on the polypeptides they encode, the two classes of clones are referred to as the a2C and a2C2α variants, respectively. From a fibroblast cDNA library, we isolated a 3.4-kb cDNA clone, F221, which was essentially identical to the a2C2α variant, except that an extra 293-bp segment was found at the end of the common region (Fig. 1A). The insertion introduced an early in-frame stop codon (Fig. 1B), resulting in an α2 chain with an 8-residue C2α' domain and an 877-bp 3'-untranslated region. This clone is referred to as the a2C2α' variant.

Intron-Exon Structure of the Carboxyl Globular Domain Coding Region—Southern blot analysis of the cosmid clone D1 indicated that a 10.5-kb HindIII fragment hybridized with all three probes (probes b, d, and f in Fig. 1A) isolated from the variable regions of the three cDNA variants, respectively (Fig. 2). Using a cDNA fragment derived from the common region of the carboxyl globular domain as a probe, a 3.8-kb HindIII genomic fragment was found to contain coding sequence for the end of the common region. Accordingly, fragments from these two regions of the D1 clone were subcloned and used to identify the intron/exon arrangements. The nucleotide sequences of all exons, the intron/exon junctions, and several small introns were determined. As shown in Figs. 2 and 3, the carboxyl globular domain of the α2 chain is encoded by six exons which were numbered beginning from the 3' end of the gene. Exon 6 (46 bp) starts with a complete codon specifying the first amino acid of this domain. It is encoded by six exons which were numbered beginning from the 3' end of the gene; the size (in base pairs) and number (in parentheses) are indicated above the exons. Exons 6-3 encode the entire common region of the carboxyl globular domain. The introns in this portion of the gene are relatively small (166-250 bp); however, the next exon is located about 2.2 kb further downstream and encodes the extra sequence present in the a2C2α' variant plus the 607-bp 3'-most sequence shared by the a2C2α and a2C2α' variants. Finally, following another intron of 2.3 kb is the last exon which encodes the entire variable region of the a2C2α variant. Most introns interrupt within codons except for the intron preceding exon 6 which occurs between two complete codons.

A consensus polyadenylation signal (AATAAA) was present in exon 1 at 12 nucleotides upstream from the polyadenylation site of the a2C2 variant (Fig. 3). However, no obvious match to this hexanucleotide consensus was found in equivalent positions of exon 2, except for the sequences TTTAATAAA and ATTTTAA at 6 and 45 nucleotides, respectively, upstream from the common polyadenylation site of the a2C2α and a2C2α' variants. Two closely related sequences, ATAAAA and ATTTTTAAA, have been shown previously to serve as signals for polyadenylation in cellular and viral genes (23, 24).

To correlate the cDNA variants with the mRNA transcripts, we examined by Northern blot hybridization human fibroblast polyadenylated RNA using probes specific for each variant. An α2(VI) collagen cDNA probe which detects mRNA of 9-10 kb was included as a control (Fig. 4, lanes 1-3). The probe unique to the a2C2α variant

![Diagram](https://example.com/diagram.png)

**FIG. 2.** Restriction map of the human α2(VI) collagen genomic clone D1 (top) and the exon map (middle) of the segment encoding the carboxyl globular domain showing alternative splicing at the 3' end and the resulting mRNAs with predicted protein domains (bottom). The entire carboxyl globular domain was encoded by 6 exons (thick bar) contained within a 3.8 kb HindIII fragment and a 10.5-kb HindIII fragment of the D1 cosmid clone. The exons are numbered from the 3' end of the gene; the size (in base pairs) and number (in parentheses) are indicated above the exons. Exons 6-3 encode the common region and exons 2 and 1 code for the variable regions. The alternative splicing of the three different mRNA species observed is indicated by the filled lines. TGA and TAA indicate in-frame translational stop codons and AATAAA represents a polyadenylation signal. Restriction enzyme sites: B, BamHI; C, CiaI; H, HindIII; P, PstI; S, SacI; X, XhoI. The abbreviation used for the protein domains are the same as in Fig. 1. 3'-UT represents noncoding sequence at the 3' end. Arrowheads (middle) denote a 1.5-kb fragment used for Northern hybridization.
Alternative Splicing of the Human a2(VI) Collagen Gene

Fig. 3. Partial nucleotide sequence of 3' end of the human a2(VI) collagen gene. The exon sequences are given in upper case letters, and the intron and flanking sequence are in lower case letters. Amino acids encoded within the exons are labeled in the one-letter code beneath the second nucleotide of each codon. The PCR primers (01-05) are underlined; three 30-40-bp segments which share 70-73% sequence homology are underlined. Downward arrow demarcates the internal splice acceptor site. Consensus and possible polyadenylation signals are boxed.

TABLE I
Splice junction sites in the carboxyl globular domain coding region of the human a2(VI) collagen gene

| Splice acceptor sites | Exons | Splice donor sites |
|-----------------------|-------|-------------------|
| 5'                  |       | 3'                |
| cctgccactggcacGALT    | Exon 6 | AGGCTgagggcagcaggcca |
| cttgccactggcacGACTG   | Exon 5  | -4CAAGtgccttcgagaagcag |
| tccctctctctctcagAGCC  | Exon 3  | -3ACAGtgccttcgagaagcag |
| tctgtctctctctcgtgAAGT | Exon 2  | 3' (end of a2C2') |
| tccctctctctctcctggAGCT | Exon 1  | 3' (end of a2C2a) |
| tccctctctctctctcctggAGCT | Exon 1  | 3' (end of a2C2a) |
| TTTAAGGGTATATTACTCT   | Exon 1  | 3' (end of a2C2a) |
| GCAAGGCTGAGGGCAAGCAG  | Exon 1  | 3' (end of a2C2a) |
| CAGCAGCTGAGGGCAAGCAG  | Exon 1  | 3' (end of a2C2a) |
| CAGCAGCTGAGGGCAAGCAG  | Exon 1  | 3' (end of a2C2a) |
| Consensus acceptor site | (t/c)3 | Consensus donor site | AAG(t/a/g) |

a Internal splice acceptor site in exon 2 (Fig. 2).

b From Ref. 26.

showed a strong hybridization to a broad band around 3.5 kb and a weak hybridization to a distinct band of about 6.0 kb (Fig. 4, lane 1), a pattern consistent with our previous findings using cDNA probe containing common region (19). A probe which recognizes both the a2C2a and a2C2a' variants showed weak hybridizations at about equal intensities to the 6.0 and 3.2 kb bands. The latter corresponds to approximately the size of the a2C2a transcript. However, a probe consisting of the inserted segment of the a2C2a' variant failed to detect mRNA transcript (Fig. 4, lane 2), even after prolonged exposure of the Northern blot (not shown). The results thus suggest that, in human fibroblasts, the a2C2 variant is most prevalent. The a2C2a variant is also present but at a significantly lower level.

Since the 6.0 kb mRNA hybridized to probes from both exon 1 and exon 2, presumably, it corresponds to incompletely processed mRNA transcripts. Therefore, one of the Northern blot filters was re-probed with a 1.5-kb PstI fragment isolated from the last intron. The probe was found to hybridize only to the 6.0-kb band (Fig. 4, lane 4). The expected size for the mRNA containing both exon 1 and exon 2, plus the last intron should be about 6.3 kb, and this corresponds well with the
Alternative Splicing of the Human α2(VI) Collagen Gene

Fig. 4. Northern blot analysis of poly(A)+ RNA from human skin fibroblasts. Each lane contains 1 μg of poly(A)+ RNA on a 1% agarose-formaldehyde gel and transferred to a nylon filter. The filters were hybridized with probes b, d, and f as depicted in Fig. 1A. Lane 1, probe b, recognizes the α2C2 variant; lane 2, probe d, recognizes both the α2C2a and α2C2a' variants; lane 3, probe f, recognizes the α2C2a' variant; lane 4, a 1.5-kb PstI fragment from the last intron as indicated in Fig. 2. In lanes 1-3, a cDNA probe for the α3(VI) collagen was included as a control. The autoradiography was performed overnight at -70°C.

The complexity of the α2(VI) mRNA transcripts was also examined by using a more sensitive method, the polymerase chain reaction, to amplify DNA fragments corresponding to each variant. A primer (01, in Fig. 1A) from the end of the constant region was paired with each of the three primers (02, 04, and 05, in Fig. 1A) derived from the variable regions of the different cDNAs. Also, primers 03 and 04 were used to amplify a region common to the α2C2a and α2C2a' variants. Single-stranded cDNA synthesized from fibroblast total RNA was amplified with each of the four primer pairs and the amplified fragments were sized on an agarose gel. As controls, parallel reactions were performed using cloned cDNAs as templates. The gel was Southern blotted and hybridized with probes specific for each variant. The predicted sizes of the amplified fragments with various primer pairs are 137 bp for the α2C2a' variant, 412 and 431 bp for the α2C2a variant, and 195, 412, and 722 bp for the α2C2a' variant. Fig. 6 shows that the expected fragments for all three variants were observed and the specificity of the amplification was confirmed by Southern hybridization. The intensity of the amplified fragments was similar; however, one may or may not reflect the relative abundance of the corresponding mRNAs since the yield of PCR products depends on both the fragment size and the primer efficiency. The fact that the α2C2a' variant can be seen using PCR amplification, but cannot be detected by Northern blot suggests that this form is present in fibroblasts but in very low abundance. It should be noted that when primers 01 and 04 were used that would generate fragments from both the α2C2a and α2C2a' variant mRNAs, only the α2C2a variant was amplified. The result could be caused by the fact that the α2C2a is more abundant than α2C2a' plus...
Alternative Splicing of the Human α2(VI) Collagen Gene

Fig. 6. PCR analysis of the α2(VI) collagen mRNA transcripts in human skin fibroblasts. PCR was performed with specific primers (depicted in Fig. 1) after first strand cDNA synthesis using total RNA isolated from human fibroblasts (lanes 1-4). The reaction products were separated on a 1% agarose gel, and the DNA bands were visualized with ethidium bromide under ultraviolet light (A). Southern blot analysis (B) was performed using the three α2(VI) collagen variant specific probes (b, d, and f depicted in Fig. 1). Controls consisted of performing PCR with the same primers using variant cDNA clone as a template (lanes 6-9). For each panel: lanes 1 and 6, primers 01 and 02, specific for the α2C2 variant; lanes 2 and 7, primers 01 and 05, specific for the α2C2a variant; lanes 3 and 8, primers 03 and 04, specific for the α2C2a and α2C2a′ variants. Lanes 4 and 9, primers 01 and 04, also specific for the α2C2a and α2C2′ variants. Lane 5, control primers A8 and A76, specific for the 3′ untranslated region of α1(I) collagen. Lane M, DNA size markers.

The α2C2a fragment is smaller and thus being preferentially amplified.

DISCUSSION

Exons Delineate the Structural Subdomains of the Carboxy Globular Domain—The results reported here demonstrate that the carboxy globular domain of α2(VI) collagen is encoded by 6 exons spanning about 7.6 kb of human genomic DNA. A striking feature of the structure of this portion of the α2(VI) collagen gene is that the beginning of the carboxy globular domain is delineated by the 5′ end of exon 6. The splice point occurs between two complete codons, specifying the last amino acid of the triple-helical domain and the first amino acid of the carboxy globular domain, respectively. This organization is in sharp contrast to the equivalent regions of all other collagen genes reported to date. In the genes for the major fibrillar collagens, types I–III, a junction exon encodes the end of the triple-helical domain, the telopeptide, and the beginning of the carboxy propeptide (see Ref. 27). Likewise, in the genes for nonfibrillar collagens, types IV, IX, X, XIII, the junction of collagenous and noncollagenous domains are always encoded by a single exon (28-31). Type VI collagen is unique among collagens in its primary structure and can be considered as a hybrid molecule of a collagen and a globular protein, since its globular domains comprise more than two thirds of the total mass and homologous globular domains are found in other noncollagenous proteins. The exon/intron organization of the 3′ end of the collagen VI gene is, therefore, compatible with the hypothesis that the gene may have originated from a recombination event involving primordial genes for a collagen and a globular protein.

Another feature of the 3′ end of the α2(VI) collagen gene is that individual structural subdomains appear to be encoded by separate exons, as has been noted in many other proteins (32). The carboxy globular domain of α2(VI) collagen consists of two repeating units that are also found in the amino and carboxy globular domains of all three chains of collagen VI. These repeats are separated from each other from the collagenous domain by short connecting segments (CS) of 25–30 residues long enriched in cysteines (Fig. 1 and Ref. 8). As shown in Fig. 2, exons 6 and 3 contain most of the coding information for each of the two short connecting segments, CS2 and CS3, respectively. The C1 repeat is encoded by two exons, whereas the C2 repeat and the alternative C2a or C2a′ domains are each coded for by a single exon (see below).

The exon structure of other genes encoding domains similar to the C1 and C2 domains of the type VI collagen have been reported (33, 34). The arrangement of exons encoding this domain vary widely. In four cases, this domain is encoded by two to five exons; in another case, two consecutive domains are encoded by a single exon. The C1 domain of collagen VI is encoded by two exons, similar to the exon structure for the two domains of the cartilage matrix protein (34). However, in contrast to the latter two domains, the internal intron in the C1 domain of collagen VI occurs within the first half of the peptide domain instead of the second half. Therefore, the exon/intron arrangements of C1 and C2 represent two new variations of the already divergent exon patterns for these homologous domains. The results suggest that although the amino acid sequences are homologous, the gene structures encoding these repeats have diverged for some time.

Alternative Splicing of the Last Two Exons Generate Multiple mRNA Transcripts Which Predict Three Protein Variants with Distinct Carboxy Termini—Cloning of the cDNAs indicated that at least three different mature mRNA transcripts are produced from the α2(VI) collagen gene. In the present study, we establish the relation between the three mRNA species. The α2C2, α2C2a, and α2C2a′ transcripts use alternative 3′ splice acceptor sites and polyadenylation sites. The most 3′ exon unique to each variant is linked to a common set of 5′ exons. Consequently, the protein variants encoded by this gene share most of the sequences, except for the carboxyl-terminal portions. In addition, Northern blot hybridization revealed a 6.0-kb mRNA transcript which retains the last intron. This RNA predicts a protein variant of the α2C2a type.

The expression of the multiple α2(VI) collagen transcripts in human skin fibroblasts indicate that the α2C2a, α2C2a′, and the 6.0-kb species are significantly less abundant than the α2C2 species. This is also in agreement with protein data showing the α2C2 variant as the only one so far detected in tissue extracts (8). Differential production of these mRNA transcripts may be controlled by the selection of either the 3′ splice acceptor sites or the specific polyadenylation sites. Inspection of the sequences of the three alternative splice acceptor sites in the α2(VI) collagen gene indicated that they do not differ significantly from each other when compared with the consensus splice acceptor sequence. On the other hand, a canonical polyadenylation signal was present in exon
1, but missing in exon 2, which encodes the two minor transcripts. This raises the possibility that utilization of the distal polyadenylation site may favor the selection of the 3' splice acceptor site in the last exon. However, we cannot exclude the possibility that the alternative splicing events are themselves controlling the choice of the cleavage and polyadenylation.

Divergence in the carboxyl termini generated from alternative splicing accompanied by utilization of alternative polyadenylation sites has been found in a number of eukaryotic genes, including genes for immunoglobulin heavy chains, myosin heavy and light chains, calcitonin, fibrinogen, and the β3 subunit of integrins (35; for a review, see Ref. 36). Alternative processing of these mRNA transcripts is often regulated in a tissue or developmental stage specific manner. In the case of the immunoglobulin heavy chain gene, the functional significance of the alternatively spliced polypeptides is well documented. Alternative utilization of an exon encoding a membrane-anchoring domain allows the cells to switch from a membrane-bound to a secreted form (see Ref. 35). The function of the different polypeptides produced by most other genes are not as clearly defined. For example, the isoforms of myosin heavy chain in Drosophila are proposed to play a role in generating myofilaments of distinct organizations (37). The alternative forms of integrin differ in the cytoplasmic domain and may alter their interaction with cytoskeletal components (36).

The significance of the carboxyl-terminal variation of type VI collagen is unknown, however, several predictions can be made based on the available structural information. Analysis of the predicted amino acid sequence of the three α2(VI) variants revealed no apparent membrane-spanning domain. It is, therefore, unlikely that the variation in the carboxyl terminus affects subcellular localization. The carboxyl globular domain of type VI collagen shares structural homology with the collagen-binding domain of von Willebrand factor, suggesting that it may interact with collagenous sequences in its vicinity. It is in this context of interest that the sequence changes remove 1 cysteine each from the α2C2a and α2C2a' variants, whereas another single cysteine directly at the carboxyl terminus is conserved in all three variants (Fig. 3). This could cause variations in disulfide bonding pattern which may influence the microfibrillar assembly of type VI collagen (8). Therefore, carboxyl-terminal variation may affect the organization of extracellular matrices containing type VI collagen.

The production of the α2(VI) chain isoforms could also constitute a mechanism to control the amount of the mature type VI collagen made by a particular cell. In the fibrillar collagens, the carboxyl propptide is thought to be crucial for the assembly of the three pro-α chains into the triple-helical molecule (38). If this is also the case for type VI collagen, the association of molecules containing distinct carboxyl termini may occur at dramatically different rates, thereby controlling the triple helix formation. It should be noted that there is no direct evidence that the α2C2a and α2C2a' forms are found in the triple-helical molecules. Alternatively, these two forms may be nonfunctional in that they are unable to be incorporated into the mature protein, which could, therefore, serve as a potential mechanism to control the amount of functional α2(VI) chain. In this regard, recent studies in Drosophila indicated that alternative splicing may be a rather common mechanism to turn the expression of the gene products on and off (39).

Alternative splicing has been reported in other collagen genes. In the α1(XIII) collagen gene, at least two exons encoding distinct parts of the collagenous region are either present or absent in the mRNA transcript (31). The physiological significance of these alternative splicing events is still unknown. The α1(IX) collagen gene, however, transcribes two mRNAs that differ by 700 bp at the 5' end in cartilage as compared to cornea, probably as a result of alternative utilization of two promoters in a tissue-specific manner (40). Similarly, the α2(I) collagen gene utilizes an alternative transcription start site in chondrocytes grown in suspension (41). The α2(VI) collagen gene is, therefore, the first example of a collagen gene in which alternative splicing gives rise to divergence in the carboxyl-terminal coding region and 3'-noncoding region. We do not yet know the temporal or spatial expression of these protein variants. It may be relevant to note that in the placental cDNA library, we found almost equal numbers of the α2C2 and α2C2a cDNA clones (8). To understand the biological significance of the carboxyl-terminal divergence in type VI collagen requires further information about the function and expression of these protein variants.

Acknowledgments—We thank Dr. Francesco Ramirez for valuable discussions and critical reading of the manuscript, Loretta Renkart for excellent technical assistance, and Sandy Parsons for typing the manuscript.

REFERENCES
1. Timpl, R., and Engel, J. (1987) in Structure and Function in Collagen Types (Mayne, R., and Burgess, R. E., eds) pp. 105-143, Academic Press, Orlando, FL.
2. Keene, D. R., Engvall, E., and Glanville, R. W. (1988) J. Cell Biol. 107, 1995-2006.
3. Aumailley, M., Mann, K., von der Mark, H., and Timpl, R. (1989) Exp. Cell Res. 181, 465-474.
4. Jander, R., Rettenmaier, J., and Glanville, R. W. (1983) Eur. J. Biochem. 133, 39-46.
5. Trueb, B., and Winterhalter, K. (1986) EMBO J. 5, 2815-2819.
6. Colombatti, A., Bonaldo, P., Ainger, K., Bresnan, G. M., and Volpin, D. (1987) J. Biol. Chem. 262, 14454-14460.
7. Chu, M.-L., Conway, D., Pan, T.-C., Baldwin, C., Mann, K., Deutmann, R., and Timpl, R. (1989) J. Biol. Chem. 264, 15801-15806.
8. Chu, M.-L., Pan, T.-C., Conway, D., Kuo, H.-J., Glanville, R. W., Timpl, R., Mann, K., and Deutmann, K. (1989) EMBO J. 8, 1939-1946.
9. Corbi, A. I., Miller, L. J., O’Connor, S., Larson, S., and Springer, T. A. (1987) EMBO J. 6, 4023-4028.
10. Titani, K., and Walsh, K. A. (1988) Trends Biochem. Sci. 13, 70-77.
11. Bentley, D. R. (1986) Biochem. J. 239, 339-345.
12. Argraves, W. S., Deak, F., Sparks, K. J., Kiss, I., and Goetinck, P. F. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 464-468.
13. Killer, E., Winterhalter, K. H., and Trueb, E. (1989) EMBO J. 8, 1073-1077.
14. Weil, D., Mattei, M. G., Passage, E., Van Cong, N., Pribula-Conway, D., Mann, K., Deutmann, R., Timpl, R., and Chu, M.-L. (1988) Am. J. Hum. Genet. 42, 435-445.
15. Vissing, H., Grosveled, F., Solomon, E., Moore, G., Lench, N., Shennan, N., and Williamson, R. (1987) Nucleic Acids Res. 15, 1363-1375.
16. Sanger, F., Nicklen, S., and Couison, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 208, 4965-4967.
17. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 166, 156-159.
18. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412.
19. Chu, M.-L., Mann, K., Deutmann, R., Pribula-Conway, D., Heu-Chen, C. C., Bernard, M. P., and Timpl, R. (1987) Eur. J. Biochem. 168, 509-517.
20. Deak, A. J., and Sharp, P. A. (1970) Proc. Natl. Acad. Sci. U. S. A. 75, 1274-1278.
21. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-500.
22. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi,
Alternative Splicing of the Human α2(Ⅵ) Collagen Gene

R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Science 239, 487–491
23. Setzer, D. R., McGrogan, M., Nunberg, J. H., and Schmike, R. T. (1980) Cell 22, 361–370.
24. Swimmer, C., and Shenk, T. (1986) Nucleic Acids Res. 13, 8053–8063
25. Brown, J. W. S. (1980) Nucleic Acids Res. 14, 9549–9559
26. Shapiro, M. B., and Senapathy, P. (1987) Nucleic Acids Res. 16, 7155–7174
27. Ramirez, F., Bernard, M., Chu, M.-L., Dickson, L., Sangiorgi, F., Well, D., de Wet, W., Juniciu, C., and Sobel, M. (1985) Annu. N. Y. Acad. Sci. 460, 117–129
28. Soininen, R., Huotari, M., Ganguly, A., Prockop, D. J., and Tryggvason, K. (1989) J. Biol. Chem. 264, 13565–13571
29. Lozano, G., Ninomiya, Y., Thompson, H., and Olsen, B. R. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 4050–4054
30. LuValle, P., Ninomiya, Y., Rosenblum, N. D., and Olsen, B. R. (1988) J. Biol. Chem. 263, 18378–18385
31. Tikka, L., Pihla-Ihami, T., Henttu, P., Prockop, D. J., and Tryggvason, K. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7491–7495
32. Patty, I. (1985) Cell 41, 657–663
33. Sadler, J. E., Shelton-Inloes, B. B., Sorensen, J. M., and Titani, K. (1986) Cold Spring Harbor Symp. Quant. Biol. 51, 515–523
34. Kiss, I., Deak, F., Holloway, R. G., Delius, H., Mebus, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., and Goetinck, P. (1980) J. Biol. Chem. 264, 8126–8134
35. Leff, S. E., and Rosenfeld, M. G. (1986) Annu. Rev. Biochem. 55, 1091–1117
36. van Kuppevelt, T. H. M. M., Languino, L. R., Gailit, J. O., Suzuki, S., and Ruoslahti, E. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 5415–5418
37. Berubein, S. I., Hansen, D. J., Decker, K. D., Wasseberg, D. R., Roche, E. S., Donady, J. J., and Emerson, C. P. (1986) Mol. Cell. Biol. 6, 2511–2519
38. Fessler, L. I., Timpl, R., and Fessler, J. H. (1981) J. Biol. Chem. 256, 2531–2537
39. Bingham, P. M., Chou, T.-B., Mims, L., and Zachar, Y. (1988) Trends Genet. 4, 134–138
40. Svoboda, K. K., Nishimura, I., Sugrue, S. P., Ninomiya, Y., and Olsen, B. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7496–7500
41. Bennett, V. D., Weiss, I. M., and Adams, S. L. (1989) J. Biol. Chem. 264, 8402–8409
Alternative splicing of the human alpha 2(VI) collagen gene generates multiple mRNA transcripts which predict three protein variants with distinct carboxyl termini.

B Saitta, D G Stokes, H Vissing, R Timpl and M L Chu

J. Biol. Chem. 1990, 265:6473-6480.