We report the identification of three new collagen VI genes at a single locus on human chromosome 3q22.1. The three new genes are COL6A4, COL6A5, and COL6A6 that encode the α4(VI), α5(VI), and α6(VI) chains. In humans, the COL6A4 gene has been disrupted by a chromosomal break. Each of the three new collagen chains contains a 336-amino acid triple helix flanked by seven N-terminal von Willebrand factor A-like domains and two (α4 and α6 chains) or three (α5 chain) C-terminal von Willebrand factor A-like domains. In humans, mRNA expression of COL6A5 is restricted to a few tissues, including lung, testis, and colon. In contrast, the COL6A6 gene is expressed in a wide range of fetal and adult tissues, including lung, kidney, liver, spleen, thymus, heart, and skeletal muscle. Antibodies to the α6(VI) chain stained the extracellular matrix of human skeletal and cardiac muscle, lung, and the territorial matrix of articular cartilage. In cell transfection and immunoprecipitation experiments, mouse α4(VI)N6-C2 chain co-assembled with endogenous α1(VI) and α2(VI) chains to form trimeric collagen VI molecules that were secreted from the cell. In contrast, α5(VI)N5-C1 and α6(VI)N6-C2 chains did not assemble with α1(VI) and α2(VI) chains and accumulated intracellularly. We conclude that the α4(VI)N6-C2 chain contains all the elements necessary for trimerization with α1(VI) and α2(VI). In summary, the discovery of three additional collagen VI chains doubles the collagen VI family and adds a layer of complexity to collagen VI assembly and function in the extracellular matrix.

Collagen VI is an extracellular component that is present in virtually all connective tissues, where it forms abundant and structurally unique microfibrils in close association with basement membranes. Collagen VI interacts with a range of ECM2 components. However, its precise role is not clearly understood. Several recent studies have suggested that collagen VI functions to anchor the basement membrane to the pericellular matrix in muscle (1–3). Other data suggest a role for collagen VI in cell signaling and cell migration (4, 5).

Three genetically distinct collagen VI chains, α1(VI), α2(VI), and α3(VI), encoded by the COL6A1, COL6A2, and COL6A3 genes were first described more than 20 years ago (6–8). The COL6A1 and COL6A2 genes are located in tandem on chromosome 21q22.3. The α1(VI) and α2(VI) chains are similar in size and domain structure. They contain a short 335- or 336-amino acid triple helix with a glycine triplet repeat motif that is characteristic of all collagens. Flanking the triple helix are domains homologous to the A-type domains found in von Willebrand factor (VWA domains). α1(VI) and α2(VI) contain one VWA domain N-terminal to the triple helix (N1) and two VWA domains on the C-terminal flank of the helix (C1 and C2). In contrast, the α3(VI) chain, encoded by the COL6A3 gene on 2q37.3, is much larger with 10 N-terminal (N1–N10), two C-terminal VWA domains (C1 and C2), and several other identifiable types of domains at the C terminus (C3–C5).

Although the assembly of collagen VI is a complex process, some of the major details have been described (6–16). The first step is monomer formation where the three chains trimerize in a 1:1:1 ratio. The exact mechanism of how this occurs is unknown, although there is evidence that the C1 domains immediately C-terminal to the triple helix are involved in initial chain selection (15, 16). These heterotrimeric monomers then associate in an anti-parallel fashion to form dimers, followed by the lateral association of dimers into tetramers. Dimers and tetramers are stabilized by disulfide bonds formed between cysteine residues present in the triple helix of all three chains (17–19). The tetramers are then secreted from the cell where they associate in an end-to-end manner to form collagen VI microfibrils. These microfibrils have a characteristic beaded appearance of globular domains separated by a rod-like triple helix with 105 nm periodicity.

Mutations in any of the three collagen VI chains lead to two types of congenital myopathies, the relatively mild Bethlem myopathy (BM) and the more severe Ullrich congenital muscular dystrophy (UCMD) (reviewed in Ref. 20). It is increasingly recognized that UCMD and BM are part of the same clinical spectrum, and the effect of a particular mutation on the production of collagen VI determines the severity. Interestingly, in a significant proportion of individuals affected with these myopathies, mutations in COL6A1, COL6A2, or COL6A3 cannot be found (21, 22) suggesting that other molecules are involved in the pathology of these diseases.

Here we report the identification and characterization of three additional collagen VI genes, COL6A4, COL6A5, and COL6A6. These genes are clustered at a single genomic locus in mammals and are located in humans on the q arm of chromo-
Three Novel Collagen VI Genes

rhesus monkey (chr 2), dog (chr 23), rat (chr 8), mouse (chr 9), platypus and lizard

some 3. The new chains encoded by these genes, α4(VI), α5(VI), and α6(VI), are most similar to the α3(VI) chain. However, evidence is presented that only the α4(VI) encompassing the N6 to C2 domains, but not the α5(VI)N5 to C1 and α6(VI)N7 to C2 chains, are competent to co-assemble with α1(VI) and α2(VI) chains. This finding suggests either that alternative collagen VI assemblies are possible or the regions critical for assembly lie outside N5 to C1 for the α5(VI) chain or N7 to C2 for the α6(VI) chain.

EXPERIMENTAL PROCEDURES

Cloning of Col6a4, Col6a5, and Col6a6 cDNA Sequences—Partial and full-length human COL6A5 and COL6A6 and mouse Col6a4 cDNAs were amplified by PCR from cDNA templates and cloned into the pCCR1 vector using the CopyControl PCR cloning kit (Epicenter Biotechnologies). Primers were designed to span at least one intron-exon boundary to prevent amplification from contaminating genomic DNA. The mouse Col6a4, Col6a5, and Col6a6 cDNAs were amplified using the mcol6a4f1 and mcol6a4r1, mcol6a4f15 and mcol6a4r5, mcol6a5f1 and mcol6a5r1, and mcol6a6f1 and mcol6a6r1 primers, respectively. For amplification of human COL6A5, COL6A4, and COL6A6 cDNAs, the COL6A5f1 and COL6A5r1, COL6A6f10, and COL6A5r4, and COL6A6f2 and COL6A6f6 primer pairs were used, respectively (supplemental Table 1). The glyceraldehyde-3-phosphate dehydrogenase (G3PDH) was used to correct for differences in the amount of template RNA.

mRNA Expression Analysis—Panels of mouse (Zyagen Laboratories) and human adult and fetal (Clontech) cDNAs were subjected to PCR using the Expand High Fidelity PCR system (Roche Applied Science) according to the manufacturer’s instructions. Primers were designed to span at least one intron-exon boundary to prevent amplification from contaminating genomic DNA. The mouse Col6a4, Col6a5, and Col6a6 cDNAs were amplified using the mcol6a4f1 and mcol6a4r1, mcol6a4f15 and mcol6a4r5, mcol6a5f1 and mcol6a5r1, and mcol6a6f1 and mcol6a6r1 primers, respectively. For amplification of human COL6A5, COL6A4, and COL6A6 cDNAs, the COL6A5f1 and COL6A5r1, COL6A6f10, and COL6A5r4, and COL6A6f2 and COL6A6f6 primer pairs were used, respectively (supplemental Table 1). The glyceraldehyde-3-phosphate dehydrogenase (G3PDH) was used to correct for differences in the amount of template RNA.

FIGURE 1. The collagen VI gene locus in vertebrates. A, schematic showing the organization of the collagen VI gene locus in rhesus monkey (Macaca mulatta), dog (Canis familiaris), rat (Rattus norvegicus), mouse (Mus musculus), platypus (Ornithorhynchus anatinus), and the green anole lizard (Anolis carolinensis). The Col6a4, Col6a5, and Col6a6 and several flanking genes (sh3bp5 and capn7 and pik3r4 and atp2c1) are represented by arrowed boxes to indicate the orientation of each gene. The three collagen VI genes are arrayed in tandem in the rhesus monkey (chromosome (chr) 2, 157.009–157.486 Mb), dog (chromosome 23, 30.494–30.950 Mb), rat (chromosome 8, 110.803–111.016 Mb), and mouse (chromosome 9, 105.601–105.999 Mb) genomes. The three genes are also present in the platypus and lizard genomes, although these genomic segments have yet to be assigned chromosomes. B, same locus in the human and chimpanzee genomes has been disrupted by a chromosome break in the COL6A4 gene. In these species, genomic DNA on the telomeric side of the break is retained on 3q, contiguous with COL6A5 and COL6A6. Note that the genes now on the centromeric side of the break, TMCC1 and TRH, are not present in the original locus and are shaded in lighter gray. Numbers above the sequence refer to the start and end of the coding regions in the human genome. Approximate position of a stop codon in the predicted COL6A4 amino acid sequence is indicated by an asterisk. C, block of genomic DNA on the centromeric side of the break has been translocated to the p arm of chromosome 3 in human and chimpanzee. Note that the two genes adjacent to the chromosome break in COL6A4, NR2C2 and ZFYVE20, are not present at the original locus and are in lighter gray. Approximate positions of two stop codons in the predicted COL6A4 amino acid sequence are indicated by asterisks.
Three Novel Collagen VI Genes

Preparation of Recombinant α5(VI) and α6(VI) and Generation of Antibodies—Polyclonal antiserum was raised in rabbits against the C1 domain of the α6(VI) chain. The C1 antigen was expressed using the His-Patch ThioFusion bacterial expression system (Invitrogen). The C1 domain of the collagen VI chain was amplified using the COL6A6 fnotI and COL6A6 rsalI primers, which were modified to include 5' NotI and 3' SalI restriction sites. Following digestion with NotI and SalI, the PCR products were cloned into the pTHIO-his(A) bacterial expression vector in-frame with the bacterial thioredoxin gene. Following transformation and selection of transformants, individual colonies were expanded and screened for the presence of recombinant protein. The recombinant α6(VI)-C1 domain was found to be in the insoluble bacterial fraction. Therefore, to prepare the protein for antibody production, the insoluble bacterial lysate was resolved on SDS-polyacrylamide gels, and the glutathione-C1 domain fusion protein was excised from the gel. The recombinant protein was electroeluted from the gel slice. This purified eluted antigen was then used to raise antiserum in rabbits (Rockland Immunochemicals, Inc.). The resulting antiserum were further affinity-purified against the recombinant antigen immobilized on a nitrocellulose membrane as described (24, 25). Serum from the same rabbit, taken prior to immunization with the α6(VI) antigen, was used as a negative control.

Immunoblot Analysis—A fetal human skeletal muscle tissue extract enriched for structural proteins was obtained commercially (Biochain, catalog number P5244171). A sample was digested in 10 units of bacterial collagenase for 1 h at 37°C in digestion buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 5 mM CaCl2). 2 μg of the undigested and collagenase-digested extract were electrophoresed on a 5% (v/v) SDS-polyacrylamide gel and transferred to an Immobilon-P polyvinylidene difluoride membrane (Millipore). The membrane was blocked in 5% milk powder in Tris-buffered saline (TBS, 150 mM NaCl, 50 mM Tris-HCl, pH 7.4) and then incubated with α6(VI) chain antisera (1:5000 dilution) overnight in antibody buffer (0.5% milk powder in TBS with 0.1% Tween 20). Following washing, the signal was developed using the Western Lightning chemiluminescence system (PerkinElmer Life Sciences) and autoradiography using X-Omat film.

Immunohistochemistry—Affinity-purified α6(VI)C1 anti-serum was used to stain frozen sections of human articular cartilage obtained from surplus allograft material and a normal adult human tissue panel (Biochain Institute). Sections were blocked with 1% bovine serum albumin in PBS and incubated with dilutions of α6(VI)-C1 antiserum in PBS for 1 h. To facilitate antibody penetration, the cartilage sections were treated with 2 mg/ml hyaluronidase (bovine, type IV; Sigma) and then treated with 0.3% (v/v) H2O2 prior to incubation with primary antibody. After blocking with 1% bovine serum albumin in three washes in antibody buffer, anti-rabbit IgG-horseradish peroxidase secondary antibody (Dako) was added at a dilution of 1:10,000 and incubated for 1 h in antibody buffer. Following washing, the signal was developed using the Western Lightning chemiluminescence system (PerkinElmer Life Sciences) and autoradiography using X-Omat film.
PBS, the sections were immunolabeled with the α6(VI)-C1 or preimmun serum in PBS for 1 h. Following washing, the sections were incubated with either goat anti-rabbit or antimouse secondary antibodies. Following a further round of washing, the color was developed using the Vectastain Elite ABC kit (Vector Laboratories). Bound antibodies were visualized using 3,3’-diaminobenzidine peroxidase substrate kit (Vector Laboratories), and sections were counterstained with hematoxylin and mounted in aqueous mounting media. As a control, a monoclonal antibody directed against the triple helical domain (dark gray) highlighted VWA domains are indicated by asterisks. The signal peptide at the N terminus is a line and numbered black box, and unique domains between C2 and C3 and at the C terminus are represented by empty boxes.

Three Novel Collagen VI Genes

**FIGURE 3. Amino acid sequence and domain structure of the α5 collagen VI chain.** A, protein sequence of α5(VI) with seven N-terminal (N1–N7) and three C-terminal (C1–C3) VWA domains (light gray) and the triple helical domain (dark gray) highlighted. Cysteines are indicated by asterisks. The signal peptide cleavage site between amino acid 18–19 (TLA-DQ) indicated by the arrow was predicted by SignalP 3.0 (29). The amino acid sequence is derived from the COL6A5 DNA sequence (supplemental Fig. S1). B, domain structure of the α5(VI) chain. VWA domains are boxed and numbered N1–N7 and C1–C3. The triple helical domain is indicated by a line between N1 and C1. The signal peptide at the N terminus is a black box, and unique domains between C2 and C3 and at the C terminus are represented by empty boxes.
RESULTS

Identification of Three Novel Collagen VI Genes—Data base searching using a VWA sequence from the α3 chain of collagen VI as a probe identified an uncharacterized sequence on human chromosome 3q22.1 that encoded two VWA domains. Close inspection of the surrounding sequence revealed that the locus is evolutionarily conserved, including the flanking genes, is also present in a reptile species (green anole lizard) indicating that the locus is evolutionarily conserved. Cysteines are indicated by asterisks. The predicted signal peptide cleavage site (VNQ-DS) is indicated by an arrow (29). The amino acid sequence is predicted from the VAL6A6 DNA sequence (supplemental Fig. S2).

These similarities in domain organization and triple helix length strongly suggest that these three collagen chains belong to the collagen VI subfamily.

New Collagen VI Gene Nomenclature—The three existing collagen VI genes/proteins are named COL6A1/α1(VI), COL6A2/α2(VI), and COL6A3/α3(VI). COL6A1 and COL6A2 are arranged in tandem on chromosome 21 and oriented in the same direction with COL6A1 being the most 5’ gene of the pair. Because all three genes at the 3’q22.1 locus are in the same orientation, we propose that the first gene in the tandem array be designated COL6A4, followed by COL6A5, and finally COL6A6. These genes encode the α4(VI), α5(VI), and α6(VI) chains, respectively. Recently, some partial sequences for these genes have appeared in the data bases, and the naming of these is consistent with the nomenclature described here.

Genomic Organization of the Human Collagen VI Locus—Identifiable orthologues of COL6A4, COL6A5, and COL6A6 are present in all mammalian species for which genome data are currently available, including macaques (rhesus monkey), canines (dog), rodents (rats and mice), and monotremes (platypus) (Fig. 1A). Interestingly, the wider chromosomal locus, including the flanking genes, is also present in a reptile species (green anole lizard) indicating that the locus is evolutionarily conserved.

Three Novel Collagen VI Genes

B

![Diagram of collagen VI gene structure](image)

**Figure 4. Amino acid sequence and domain structure of α6(VI).** A, Protein sequence of α6(VI) with seven N-terminal (N1–N7) and two C-terminal (C1 and C2) VWA domains (light grey) and the triple helical domain (dark grey) highlighted. Cysteines are indicated by asterisks. The predicted signal peptide cleavage site (VNQ-DS) is indicated by the arrow (29). The amino acid sequence is predicted from the COL6A6 DNA sequence (supplemental Fig. S2). B, Domain structure of α6(VI). VWA domains are boxed and numbered N1–N7, C1, and C2. The helix is a line between N1 and C1 VWA domains. The predicted variant peptide at the N terminus is in black and a unique C-terminal domain is empty.

| N7 | N6 | N5 | N4 | N3 | N2 | N1 | C1 | C2 |
|----|----|----|----|----|----|----|----|----|

| α6(VI) |

---

These similarities in domain organization and triple helix length strongly suggest that these three collagen chains belong to the collagen VI subfamily.
Three Novel Collagen VI Genes

First, the α3, α4, α5, and α6 chains have a Cys residue at amino acid 50 of the helix in contrast to the α1 and α2 chains that have a Cys residue at a different relative position, amino acid 89. Second, the α4, α5, and α6 chains have an interruption in the triple helix between amino acids 125 and 128, which is in the same position as in the α3 chain. The triple helix domains from the α1 and α2 chains do not have this interruption. These sequence features indicate that the three new chains are most similar to the α3(VI) chain. This is supported by a phylogenetic analysis of the triple helical domains based on the alignment shown in Fig. 2 (supplemental Fig. S4). The analysis clusters the α3, α4, α5, and α6 chains in a separate group from the α1 and α2 chains.

The α5(VI) chain has several imperfections in the Gly-X-Y repeat motif, where the Gly residue is substituted by another amino acid at amino acids 13, 153, and 306 in the α5(VI) triple helix (Fig. 2). Similarly, the α6(VI) helical domain has several imperfections in the Gly-X-Y repeat motif at amino acids 139, 156, 159, 306, and 315 in the α6(VI) triple helix. Neither the α3(VI) nor α4(VI) chains have Gly-X-Y imperfections. Each of the new chains contains a single RGD cell-binding signal.

Overall, the human COL6A5 gene is composed of 39 exons with a coding region 7836 bp in size (supplemental Fig. S1). The gene encodes a protein of 2611 amino acids (Fig. 3) with a predicted molecular mass of 287 kDa. The N terminus is predicted to contain a signal peptide cleavage site between amino acids 18 and 19 (VNQ-DS) to give a molecular mass for the secreted protein of 245 kDa. Following the signal peptide are seven VWA domains and a 336-amino acid triple helix, and then three more VWA domains. On either side of the C3 domain are regions that are 134 and 129 amino acids in size. The human COL6A4 and COL6A6 mRNA in human tissues. cDNA from 25 human adult and fetal tissues were amplified using oligonucleotide primers specific for COL6A4 (145 bp) or COL6A4 (1248 bp). PCR products were mixed for electrophoresis to illustrate relative levels of expression for each gene. cDNAs are fetal (f) brain (lane 1), adult (a) brain (lane 2), fetal lung (lane 3), adult lung (lane 4), fetal liver (lane 5), adult liver (lane 6), fetal kidney (lane 7), adult kidney (lane 8), fetal heart (lane 9), adult heart (lane 10), fetal spleen (lane 11), adult spleen (lane 12), fetal thymus (lane 13), adult thymus (lane 14), fetal skeletal muscle (lane 15), adult skeletal muscle (lane 16), placenta (lane 17), pancreas (lane 18), small intestine (lane 19), colon (lane 20), prostate (lane 21), testis (lane 22), ovary (lane 23), uterus (lane 24) and peripheral blood lymphocytes (lane 25). Lane 26 is a no-template negative control.

Three Novel Collagen VI Genes

First, the α3, α4, α5, and α6 chains have a Cys residue at amino acid 50 of the helix in contrast to the α1 and α2 chains that have a Cys residue at a different relative position, amino acid 89. Second, the α4, α5, and α6 chains have an interruption in the triple helix between amino acids 125 and 128, which is in the same position as in the α3 chain. The triple helix domains from the α1 and α2 chains do not have this interruption. These sequence features indicate that the three new chains are most similar to the α3(VI) chain. This is supported by a phylogenetic analysis of the triple helical domains based on the alignment shown in Fig. 2 (supplemental Fig. S4). The analysis clusters the α3, α4, α5, and α6 chains in a separate group from the α1 and α2 chains.

The α5(VI) chain has several imperfections in the Gly-X-Y repeat motif, where the Gly residue is substituted by another amino acid at amino acids 13, 153, and 306 in the α5(VI) triple helix (Fig. 2). Similarly, the α6(VI) helical domain has several imperfections in the Gly-X-Y repeat motif at amino acids 139, 156, 159, 306, and 315 in the α6(VI) triple helix. Neither the α3(VI) nor α4(VI) chains have Gly-X-Y imperfections. Each of the new chains contains a single RGD cell-binding signal.

Overall, the human COL6A5 gene is composed of 39 exons with a coding region 7836 bp in size (supplemental Fig. S1). The gene encodes a protein of 2611 amino acids (Fig. 3) with a predicted molecular mass of 287 kDa. The N terminus is predicted to contain a signal peptide cleavage site between amino acids 18 and 19 (VNQ-DS) to give a molecular mass for the secreted protein of 245 kDa. Following the signal peptide are seven VWA domains and a 336-amino acid triple helix, and then three more VWA domains. On either side of the C3 domain are regions that are 134 and 129 amino acids in size that do not show homology with any other protein.

The COL6A6 gene has 36 exons and a coding region of 6789 bp (supplemental Fig. S2), which codes for a 2262-amino acid protein (Fig. 4). The signal peptide is predicted to be cleaved between amino acids 18 and 19 (VNQ-DS) to give a molecular mass for the secreted protein of 245 kDa. Following the signal peptide are seven VWA domains and a 336-amino acid triple helix and then two additional VWA domains. As with the α5(VI) protein, there is a region following C2 of 91 amino acids that shows no homology with other proteins.

Because the human COL6A4 gene is disrupted, we cloned and sequenced the mouse Col6a4 cDNA sequence (supplemental Fig. S3). The mouse Col6a4 gene has an open reading frame...
Three Novel Collagen VI Genes

of 6930 bp encoding a 2309-amino acid protein with a predicted molecular mass of 248 kDa. As with the human α5(VI) and α6(VI) chains, the α4(VI) chain has seven VWA domains followed by a collagen triple helix and two C-terminal VWA domains.

Expression of COL6A5 and COL6A6 mRNA—RT-PCR was conducted to determine the distribution of Col6a4, Col6a5/Col6A5, and Col6a6/COL6A6 in mouse and human tissues. In mouse, Col6a4 mRNA was present in uterus and ovary as faint bands but not in any other tissues examined (Fig. 5A). In contrast, mRNA for Col6a5 and Col6a6 was present in a much wider range of tissues. Col6a5 was expressed in all tissues examined except prostate, although it was expressed only weakly in muscle and trachea. Col6a6 mRNA was present in all tissues except prostate and expressed only weakly in bone, small intestine, and placenta.

In human tissues, mRNA for COL6A5 was only present at high levels in lung (adult) and testis, and at lower levels in small intestine and colon (Fig. 5B). This is in contrast to the more widespread expression pattern in the mouse. All other tissues were negative for COL6A5 mRNA, including fetal lung. To confirm the low overall expression of COL6A5, the PCRs were repeated with a different primers set with identical results (data not shown). COL6A6 mRNA was found to be expressed more widely with high levels of expression in fetal and adult lung, heart, and spleen. In kidney, thymus, and skeletal muscle, there was more mRNA expression in fetal compared with adult. Conversely, in adult liver mRNA levels were higher than in fetal liver. Several other tissues expressed COL6A6 mRNA at relatively low levels including pancreas, testis, and uterus.

Expression of α6(VI) in Human Tissues—To assess the distribution of α6(VI) chains in human tissues, polyclonal antiserum was raised against the C1 domain α6(VI). The crude antiserum was then affinity-purified using the recombinant antigen. The affinity-purified antiserum was used in an immunoblotting experiment on a human fetal muscle extract (Fig. 6). The antiserum recognized a single band at 240 kDa corresponding to the molecular weight of the α6(VI) chain. The band was digested with collagenase treatment confirming that it is a collagen. The antibody was then used to stain a range of human tissues (Fig. 7). The α6(VI) chain was present in the extracellular matrix of kidney (not shown), skeletal and cardiac muscle, lung, and blood vessels (Fig. 7, C, G, K, and Q). Lower levels of α6(VI) chain were present in pancreas and spleen (Fig. 7, R and T). Notably, significant ECM staining for α6(VI) protein was absent from liver (Fig. 7O). An α3(VI) monoclonal antibody detected extracellular collagen VI in several tissues assayed, including skeletal muscle (Fig. 7A), cardiac muscle (E), lung (I), and liver (M), reflecting the wide distribution of collagen VI. The absence of significant ECM staining for the α6(VI) chain in liver, a tissue that is rich in collagen VI (compare Fig. 7M with Fig. 7O), indicates that the affinity-purified α6(VI) antibody does not cross-react with other collagen VI chains.

Collagen VI is known to be a component of the chondrocyte pericellular matrix (30, 31). Because COL6A6 mRNA was shown to be present in mouse articular cartilage (Fig. 5A), the expression of collagen VI in human articular cartilage was assessed (Fig. 8). Pericellular staining for collagen VI is clearly shown in Fig. 8, panels F and H, where human articular cartilage was immunostained with two collagen VI antibodies. In contrast to these collagen VI antibodies is the staining pattern of the α6(VI)-specific antibody, which extends significantly farther into the territorial matrix with little or no discrete pericellular staining. This can be seen in the superficial layer but is especially clear in the deeper zone.

Immunogold electron microscopy analysis also confirmed that α6(VI) chains are present in human articular cartilage (Fig. 8B). Here, antibodies to the α6(VI) chain co-localize with extracellular material that contains collagen VI, demonstrating that the labeled material is part of the same suprastructures.

Assembly of α4(VI), α5(VI), and α6(VI) in Transfected Cells—To assess whether the three new collagen VI chains are competent to assemble with α1(VI) and α2(VI) chains, transfection studies in SaOs-2 osteosarcoma cells were conducted (Fig. 9). These cells synthesize α1 and α2 chains but not any other collagen VI chains, including COL6A5 and COL6A5 (data not shown). Cells were transfected with cDNAs for mouse Col6a4 (containing domains N6–C2), human COL6A5 (domains N5–C1), or human COL6A6 (domains N7–C2). Following selection in hygromycin, the cultures were labeled with [35S]methionine, and the labeled material in the cell layer and medium fractions were immunoprecipitated with an anti-His6 monoclonal antibody (Fig. 9A) or a commercially available collagen VI antibody (Fig. 9B). Bands corresponding to the molecular weights (~230–240 kDa) of the α4(VI)N6–C2, α5(VI)N5–C1, and α6(VI)N7–C2 chains were present in the cell fractions of their respective transfections (see Fig. 9A, lanes 3, 5, and 6). No band at this molecular weight is present in the cell layer of untransfected cells (Fig. 9A, lane 1). This confirms that the cells synthesize all three chains from their respective transfected cDNAs. The α4(VI)N6–C2 chain was present in the medium fraction (Fig. 9A, lane 4) indicating that it is secreted and competent to co-assemble with α1(VI) and α2(VI). This was con-
firmed in the reverse experiment where the metabolically labeled cultures were immunoprecipitated with antiserum that recognizes the human α1(VI) and α2(VI) chains (Fig. 9B). Again, only in the cells expressing the α4(VI)N6-C2 chain was a band present in the transfected cell medium. The detection of secreted α4(VI)N6-C2 chain when immunoprecipitating with α1/α2 antibody is good evidence that the α4 chain co-assembles with these two chains. Conversely, the failure to detect α5(VI)N6-C1 and α6(VI)N7-C2 is surprising and suggests that these chains are not competent to co-assemble with α1(VI) and α2(VI) chains in this system. Longer exposure of the gels failed to detect any secreted α5(VI)N6-C1 and α6(VI)N7-C2 chains in the media fractions. To confirm that the secreted α4(VI) chain is triple helical, immunoprecipitated sample from the medium of cells transfected with His6-Col6a4N6-C2 cDNA was resolved under nonreducing conditions (Fig. 9C). The 240-kDa α4(VI) chain migrates at the top of the gel in the absence of reduction. The band corresponding to the co-precipitated α1 and α2, which is present with reduction chain, also disappears in the nonreduced sample. Together these data suggest that α1, α2, and α4 chains co-assemble into disulfide-bonded triple helical collagen VI heterotrimers that are secreted from the cell.

DISCUSSION

The discovery of three additional chains, α4(VI), α5(VI), and α6(VI), doubles the size of the collagen VI family from three to six members. A schematic, updated to show the domain structures of the three existing (α1, α2, and α3) and three new (α4, α5, and α6) human collagen VI chains, is presented in Fig. 10. Overall, the domain organization of the collagen VI chains is similar with each having a short triple helix flanked by a variable number of VWA domains. The α1 and α2 chains are smallest with just one N- and two C-terminal VWA domains. The α3 chain is largest with up to 10 N- and 2 C-terminal VWA domains as well as several domains not related to VWA domains at the C terminus. The new α4, α5, and α6 chains are closer in size to the α3 chain with seven N-terminal VWA domains and two or three C-terminal VWA domains. Several features within the triple helix domain suggest that the three new chains are more closely related to the α3(VI) chain than α1(VI) and α2(VI). The α3, α4, α5, and α6 chains have the same interruption in the Gly-X-Y repeat motif in the triple helix, which is not present in the α1 and α2 chains. In addition, α3(VI) and the three new collagen VI chains contain a single cysteine at amino acid 50 of the triple helix (Fig. 2). In contrast, the α1(VI) and α2(VI) chains have a helix cysteine at a different relative position at position 89. Analysis of disease-causing mutations in the collagen VI chains has demonstrated that these cysteine residues are crucial for forming interchain disulfide bonds and stabilizing collagen VI assembly intermediates (10, 17–19, 32, 33). These sequence features suggest that the COL6A3, COL6A4, COL6A5, and COL6A6 genes share a common evolutionary history.

The collagen VI locus at 3p22.1 is evolutionarily conserved and predates the split of the reptile lineage, although in humans the COL6A4 gene is disrupted by a chromosomal break. Rhesus monkeys appear to possess three intact collagen VI genes at this
locus, but the human and chimpanzee genomes only have two intact genes, COL6A5 and COL6A6. Therefore, the chromosome translocation that disrupted the COL6A4 gene must have occurred following the split of the hominoids (apes and humans) from the old world monkey lineage (macaques, gibbons, and vervets) estimated to be 30 million years ago (34). The break occurred in the domain that encodes the triple helix and distributed the 5′/H11032 end of COL6A4 to the p arm of chromosome 3. This is a major disruption to the gene and is highly likely to render the gene functionally inactive, although a careful analysis will need to be conducted to exclude the possibility that transcripts derived from the split gene have functional significance. This is a possibility because such a recent chromosome rearrangement may have left the regulatory regions of the 5′ end of COL6A4 to the p arm of chromosome 3. This is a major disruption to the gene and is highly likely to render the gene functionally inactive, although a careful analysis will need to be conducted to exclude the possibility that transcripts derived from the split gene have functional significance. This is a possibility because such a recent chromosome rearrangement may have left the regulatory regions of the 5′ end of the gene intact. If a partial protein for the α4(VI) chain is produced, it may exert a dominant negative effect on collagen VI function. There are 11 ESTs for the 5′ half of the gene suggesting that this portion may be transcriptionally active.

In mouse, Col6a5 and Col6a6 are transcribed in a wide range of tissues. However, expression of Col6a4 is confined to a few tissues, including uterus and ovary. In humans, there is also widespread expression of COL6A6 but restricted expression of COL6A5. The reason for the difference in expression of COL6A5/COL6a5 between human and mouse is not known. One possibility is that the chromosome break altered the genomic context of the locus such that trans-acting regulatory elements required for widespread expression of COL6A5 were removed or deleted. Nonetheless, there are multiple tissues in humans and mouse where at least one of the new chains is expressed.

With the existence of six chains the assembly of collagen VI is more complex than previously thought and implies that a chain selection mechanism must exist to distinguish between chains in cells that express more than three chains. The trimerization of the fibrillar collagens and collagen IV, which also has six chains, has been well studied (reviewed by Khoshnoodi et al. (35)). The common theme is that initial chain selection involves elements C-terminal to the triple helix, and this appears to be the case for the collagen VI chains as well (12, 15, 16). However, the motifs responsible for chain selection in collagen VI may be different because there is no clear C-terminal domain homology between the collagen IV, collagen VI, and the fibrillar collagen subfamilies.
Three Novel Collagen VI Genes

To investigate the ability of the new chains to co-assemble with α1(VI) and α2(VI), transfection and immunoprecipitation experiments in SaOs-2 osteosarcoma cells were conducted. SaOs-2 cells express mRNA for COL6A1 and COL6A2 but not COL6A3, COL6A5, or COL6A6 (see Refs. 11, 14 and data not shown), making them an ideal cell line to assess whether these chains assemble into collagen VI molecules. α1(VI) and α2(VI) are not competent to assemble into heterotrimeric collagen VI molecules alone and require a third chain for collagen VI assembly (11, 36). This allowed us to test the model that the new chains substitute for α3(VI) and co-assemble with α1(VI) and α2(VI). However, this model may be an oversimplification of the in vivo situation because although we show that the mouse α4(VI)N6-C2 chain can co-assemble with α1 and α2, the α5N5-C1 and α6N7-C2 chains do not associate with α1 and α2 and are retained inside the cell. Both chains contain the C1 domain, which is reported to be critical for collagen VI chain association (12, 15, 16). However, the COL6A5N5-C1 cDNA used in our transfection study lacked the exon for the C2 domain, and we cannot exclude the possibility that this domain is important for α5(VI) co-assembly with α1(VI) and α2(VI). Overall, these data indicate that the α5(VI) and α6(VI) chains are not competent to assemble with α1 and α2 chains and suggest that these chains assemble into collagen VI heterotrimers that are different from α1-α2-α3 and α1-α2-α4 trimers. Future studies will examine the possibility of alternative collagen VI chain assemblies that do not require α1(VI) and/or α2(VI).

Interestingly, the distribution of α6(VI) in cartilage is different from that of the α3(VI) chain. The α3(VI) chain is concentrated around the cell in the pericellular matrix, and the α6(VI) chain is predominantly in the territorial matrix, although there is overlap with the pericellular matrix (see Fig. 8B). This finding raises the possibility that not all collagen VI microfibrils contain α3(VI) chains. This suggests that different collagen VI macromolecular structures may exist.

A recent report mapped the collagen XXIX gene, COL29A1, to human chromosome 3q22.1 and described its genetic asso-
Three Novel Collagen VI Genes

citation with atopic dermatitis (37). The COL29A1 gene appears to be the COL6A5 sequence reported here. Based on sequence similarities with the collagen VI genes, we argue that the COL29A1 gene is not a new collagen subtype but belongs to the collagen VI family.

Finally, in ~40% of individuals with Bethlem myopathy and UCMD, mutations are not present in the COL6A1, COL6A2, or COL6A3 genes (21, 22). Our finding that COL6A6 is expressed in skeletal muscle raises the exciting possibility that mutations in this gene are involved in the pathophysiology of these and possibly other, related, congenital muscular dystrophies.

Acknowledgment—We thank Gerburg Pentrup for excellent technical help.

REFERENCES

1. Ishikawa, H., Sugie, K., Murayama, K., Awaya, A., Suzuki, Y., Noguchi, S., Hayashi, Y. K., Nonaka, I. and Nishino, I. (2004) Neurology 62, 620–623
2. Ishikawa, H., Sugie, K., Murayama, K., Ito, M., Minami, N., Nishino, I. and Nonaka, I. (2002) Neurology 59, 920–923
3. Niitaya, T., Higuchi, I., Suehara, M., Hashiguchi, T., Shiraiishi, T., Nakagawa, M., Arimura, K., Maruyama, I. and Osame, M. (2002) Acta Neuro-pathol. 104, 67–71
4. Jimenez-Mallebrera, C., Brown, S. C., Sewry, C. A., and Muntoni, F. (2005) Cell. Mol. Life Sci. 62, 809–823
5. Irwin, W. A., Bergamin, N., Sabattelli, P., Reggiani, C., Megighian, A., Merlini, L., Braghetta, P., Columbaro, M., Volpin, D., Bressan, G. M., Bernardi, P., and Bonaldo, P. (2003) Nat. Genet. 35, 367–371
6. Chu, M. L., Mann, K., Deutzmann, R., Pribula-Conway, D., Hsu-Chen, J. C., Bernard, M. P., and Timpl, R. (1987) Eur. J. Biochem. 168, 309–317
7. 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
8. Chu, M. L., Zhang, R. Z., Pan, T. C., Stokes, D., Conway, D., Kuo, H. J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R., and Timpl, R. (1990) EMBO J. 9, 385–393
9. Lamande, S. R., Morgelin, M., Selan, C., Jobsis, G. J., Baas, F., and Bateman, J. F. (1999) J. Biol. Chem. 274, 21817–21824
10. Lamande, S. R., Shields, K. A., Kornberg, A. J., Shields, L. K., and Bateman, J. F. (1998) J. Biol. Chem. 273, 7423–7430
11. Ball, S., Bella, J., Kielty, C., and Shuttleworth, A. (2003) J. Biol. Chem. 278, 15326–15332
12. Ball, S. C., Zhang, R. Z., Sudano, D. G., Marie, S. K., Bonnemann, C. G., and Chu, M. L. (2003) Am. J. Hum. Genet. 73, 355–369
13. Fitzgerald, J., Morgelin, M., Selan, C., Wiberg, C., Keene, D. R., Lamande, S. R., and Bateman, J. F. (2001) J. Biol. Chem. 276, 187–193
14. Lamande, S. R., Morgelin, M., Adams, N. E., Selan, C., and Allen, J. M. (2006) J. Biol. Chem. 281, 16607–16614
15. Ball, S. G., Baldock, C., Kielty, C. M., and Shuttleworth, C. A. (2001) J. Biol. Chem. 276, 7422–7430
16. Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E., and Engel, J. (1983) Biochem. J. 211, 303–311
17. Odermatt, E., Risteli, J., van Delden, V., and Timpl, R. (1983) Biochem. J. 211, 295–302
18. Lampe, A. K., and Bushby, K. M. (2005) J. Med. Genet. 42, 673–685
19. Lucioli, S., Giusti, B., Mercuri, E., Vanegas, O. C., Lucarini, L., Pietroni, V., Uritizberea, A., Ben Yau, R., de Visser, M., van der Kooi, A. J., Bonne mann, C., Iannaccone, S. T., Merlini, L., Bushby, K., Muntoni, F., Bertini, E., Chu, M. L., and Pepe, G. (2005) Neurology 64, 1931–1937
20. Lampe, A. K., Dunn, D. M., von Niederhausern, A. C., Hamil, C., Aoyagi, A., Laval, S. H., Marie, S. K., Chu, M. L., Swo boda, K., Muntoni, F., Bonnemann, C. G., Flanigan, K. M., Bushby, K. M., and Weiss, R. B. (2005) J. Med. Genet. 42, 108–120
21. Lamande, S. R., and Bateman, J. F. (2001) J. Biol. Chem. 276, 15326–15332
22. Lamande, S. R., Morgelin, M., Adams, N. E., Selan, C., and Allen, J. M. (2005) Am. J. Hum. Genet. 787–791
23. Bengtsson, E., Aspberg, A., Heinegard, D., Sommarin, Y., and Spillmann, D. (2000) J. Biol. Chem. 275, 40695–40702
24. Allen, J. M., Bateman, J. F., Hansen, U., Wilson, R., Bruckner, P., Owens, R. T., Sasaki, T., Timpl, R., and Fitzgerald, J. (2006) J. Biol. Chem. 281, 7341–7349
25. Michalczyzk, A. A., Allen, J., Blomeley, R. C., and Ackland, M. L. (2002) Biochem. J. 364, 105–113
26. Kassner, A., Hansen, U., Miosge, N., Reinhardt, D. P., Aigner, T., Bruck ner-Tuderman, L., Bruckner, P., and Grassel, S. (2003) Matrix Biol. 22, 131–143
27. Fitzgerald, J., Ting, S. T., and Bateman, J. F. (2002) FEBS Lett. 517, 61–66
28. Donoghue, P. C., and Benton, M. J. (2007) Trends Ecol. Evol. 22, 424–431
29. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) J. Mol. Biol. 340, 783–795
30. Sherwin, A. F., Carter, D. H., Poole, C. A., Hoyland, J. A., and Ayad, S. (1999) Histochem. J. 31, 623–632
31. Roberts, S., Ayad, S., and Menge, P. J. (1991) Ann. Rheum. Dis. 50, 787–791
32. Baker, N. L., Morgelin, M., Peat, R., Goemans, N., North, K. N., Bateman, J. F., and Lamande, S. R. (2005) Hum. Mol. Genet. 14, 279–293
33. Timpl, R., and Chu, M.-L. (1994) in Extracellular Matrix Assembly and Structure (Yurchenco, P. D., Birk, D., and Mecham, R. P., eds) pp. 207–242, Academic Press, Orlando
34. Steiper, M. E., and Young, N. M. (2006) Trends Ecol. Evol. 21, 424–431
35. Khoshnoodi, J., Cartailler, J. P., Alvares, K., Veis, A., and Hudson, B. G. (2006) J. Biol. Chem. 281, 38117–38121
36. Colombatti, A., Mucignat, M. T., and Bonaldo, P. (1995) J. Biol. Chem. 270, 13105–13111
37. Soderhall, C., Marenholz, I., Kerscher, T., Rusc hendorf, F., Esparza-Gor dillo, J., Worm, M., Gruber, C., Mayr, G., Albrecht, M., Rohde, K., Schulz, H., Wahn, U., Hubner, N., and Lee, Y. L. (2007) Plos Biol. 5, e242