Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 940–944, February 1987
Biochemistry

Partial characterization of a low molecular weight human collagen that undergoes alternative splicing

(collagen cDNAs/short-chain collagen)

TAINA PIHLAJANIEMI*†, RAILI MYLLYLÄ*, JEROME SEYER‡, MARKKU KURKINEN§, AND DARWIN J. PROCKOP¶

*Collagen Research Unit, Department of Medical Biochemistry, University of Oulu, Oulu, Finland; ‡Veterans Administration Medical Center and Department of Medicine and Biochemistry, University of Tennessee, Memphis, TN 38104; and Departments of §Medicine and ¶Biochemistry, University of Medicine and Dentistry of New Jersey–Rutgers Medical School, Piscataway, NJ 08854

Communicated by Paul Doty, October 17, 1986

ABSTRACT **A cDNA library prepared from RNA isolated from a cultured human tumor cell line, HT-1080, was screened with a mouse cDNA clone coding for part of the -Gly-Xaa-Yaa- domain of the α2(IV) collagen chain. Four overlapping cDNA clones were characterized that coded for a low molecular weight human collagen. The cDNA clones did not, however, code for the short-chain collagens, types IX and X. The amino acid sequences derived from the clones resembled type IV collagen in that there were short interruptions in the repeating -Gly-Xaa-Yaa- sequence. The noncollagenous, carboxyl-terminal domain was, however, much shorter and contained only 18 amino acid residues. Interestingly, one of the cDNA clones contained an additional 36 nucleotides not found in an overlapping clone. The 36 nucleotides encoded four -Gly-Xaa-Yaa- repeats without changing the reading frame. Nuclease S1 mapping demonstrated that the difference between the clones was due to existence of two different mRNAs. A synthetic 24-residue peptide corresponding to the last two -Gly-Xaa-Yaa- triplets and the entire carboxyl-terminal domain was used to generate polyclonal antibodies. Electrophoretic transfer blot analysis of HT-1080 cells and normal human skin fibroblasts identified two polypeptides, M_r 67,000 and M_r 62,000, that were sensitive to bacterial collagenase.**

The collagens are major structural components of the extracellular matrix. In vertebrates they consist of 10 or more genetically distinct protein types and at least 22 genes code for their constituent α chains (1–3). The fibrillar collagen types I, II, III, and V are synthesized as precursor proα chains that include a central collagen domain of about 1000 amino acids with a repeating sequence of -Gly-Xaa-Yaa- triplets. In addition, the proα chains contain amino- and carboxyl-terminal propeptides that are cleaved after the chains are assembled into procollagen trimers. Type IV collagen is specific to basement membranes and forms a network-like structure. In contrast to fibrillar collagens, type IV collagen has a somewhat larger -Gly-Xaa-Yaa- domain that is interrupted at several sites (1–3).

Recently, several additional nonfibrillar collagens have been identified in various connective tissues. In type VI collagen, the collagen domain comprises only one-third of the protein and large globular domains are present at each end (4). The long-chain collagen or type VII collagen contains an extended triple helical domain (5). The size of type VIII collagen polypeptides is presently controversial. An "interrupted helix" (6) or a "cassette" (7) model describes type VIII collagen as composed of three chains of about M_r 180,000, arranged in a predominantly helical structure, but with interruptions in the helix at approximately one-third and two-thirds the length of the molecule. Furthermore, a recent work by Kapoor *et al.* (8) describes isolation of two structurally different pepsin-resistant M_r 50,000 polypeptides that were derived from a larger form of type VIII collagen. In an alternative model by Benya and Padilla (9), however, type VIII collagen is composed of three identical α1 chains of M_r 61,000, arranged in a predominantly helical structure, with nonhelical domains at each end. This structure is stabilized in the presence of $NaDodSO_4$ by strong noncovalent acid-labile interactions and covalent cross-links to produce γ and β chains of M_r 194,000 and 124,000. Type IX (10) and X collagen (11) polypeptides are about half the size of the proα chains of types I, II, and III procollagens. In addition, several other low molecular weight collagens have been isolated and partially characterized (12–14), and collagenous sequences are also found in several other proteins such as C1q (3), acetylcholinesterase (3), and conglutinin (15).

Here, we describe nucleotide and amino acid sequences of a low molecular weight human collagen as derived from four cDNA clones. The collagen synthesized by the cultured human tumor HT-1080 cells and human skin fibroblasts was studied by using polyclonal antibodies to a synthetic peptide derived from the cDNA sequences.

MATERIALS AND METHODS

Isolation of cDNA Clones and DNA Sequencing. The clones were isolated from a cDNA library of human tumor cell line HT-1080 that was previously prepared to isolate cDNAs for α1(IV) collagen (16). The library was screened with the mouse cDNA clone pE18 [1100 base pairs (bp)] encoding part of the -Gly-Xaa-Yaa- domain of α2(IV) collagen (17) using conditions described previously (16). The cDNA sequences were determined by the Maxam–Gilbert procedure (18).

RNA Transfer Blot Analysis and Nuclease S1 Protection Experiments. An oligo(dT)-cellulose affinity column (19) was used to extract poly(A)$^+$ RNA from HT-1080 cells and a locally established normal human skin fibroblast cell line. Electrophoresis and blotting of the RNA was performed as described (16). The nuclease S1 protection experiments were performed as described (19) except a double-stranded probe was used. To prepare a probe from recombinant plasmid HT-125, the 540-bp *Pst* I/*Pst* I insert was isolated and restricted with *Sau*3A. The antisense strand of the insert DNA was 3′ end-labeled at the single *Sau*3A site by addition of [α-^{32}P]dATP (3000 Ci/mmol; 1 Ci = 37 GBq) using the Klenow fragment of DNA polymerase I. The double-stranded 530-bp *Sau*3A/*Pst* I fragment was then isolated from a nondenaturing 5% polyacrylamide gel. Three nanograms of ^{32}P-labeled probe and 8 μg of poly(A)$^+$ RNA were used for nuclease S1 digestion at pH 4.8, and the digestion products

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed at: Department of Medical Biochemistry, University of Oulu, Kajaanintie 52 A, SF-90220 Oulu, Finland.

were examined by electrophoresis on a 5% polyacrylamide sequencing gel as described (19).

Preparation of the Synthetic Peptide and Antiserum Against It. The peptide was chemically synthesized by the solid-phase procedure of Merrifield (20) with the aid of an automatic Beckman model 990 peptide synthesizer. The completed peptide was cleaved from the resin by treatment with liquid HCl at 0°C and purified further by gel filtration through Sephadex G-25 and reverse-phase high-pressure liquid chromatography. Approximately 500 μg of the synthetic peptide was conjugated to 7.5 mg of hemocyanin with glutaraldehyde (21) and emulsified in Freund's complete adjuvant prior to intradermal injection into rabbits; booster injections were given 2 and 5 weeks later using Freund's incomplete adjuvant.

Electrophoretic Transfer Blotting of Cellular and Medium Proteins. HT-1080 cells and normal human skin fibroblasts were grown in 100-mm plates under standard conditions in Dulbecco's modified Eagle's minimal essential medium containing 10% newborn calf serum and 50 μg of ascorbate per ml. At near confluency, the cell layer and medium proteins were processed separately as described (22) with some modifications. The cell layer from one plate was scraped into 0.5 ml of homogenizing buffer (22), the cells were homogenized with a Teflon/glass homogenizer, and the sample was centrifuged at 15,000 × *g* for 10 min. Half of the supernatant was prepared for electrophoresis by adding sample buffer to give a final concentration of 2% $NaDodSO_4$, 10% glycerol, 0.5 M urea, 0.001% bromphenol blue, and 0.125 M Tris·HCl (pH 6.8) at room temperature and boiled for 5 min. The other half was digested with pepsin (Boehringer Mannheim) and prepared for electrophoresis as described (22) except 50 μg of pepsin per ml was used.

For examination of cellular proteins after bacterial collagenase digestion, *p*-aminobenzamidine, phenylmethylsulfonyl fluoride, and EDTA were omitted from the homogenizing buffer (22). Half of the cell extract was prepared for electrophoresis as described above. $CaCl_2$ was added to the other half to give a final concentration of 10 mM and 0.18 mg of bacterial collagenase per ml (Sigma, type VII) was added. The sample was then digested at 37°C for 4 hr. The reaction was stopped by adding sample buffer and was boiled for 5 min. Medium proteins were precipitated by ammonium sulfate (22) and the precipitate was suspended in sample buffer prior to gel electrophoresis.

Gel electrophoresis was performed by using 8% polyacrylamide gels containing 0.5 M urea and 0.1% $NaDodSO_4$ (23). The fractionated proteins were then blotted on nitrocellulose filters (BA85, Schleicher & Schuell) as described (21). The sheets were stained with heparin/toluidine and, after cutting into strips, destained as described (21). The nitrocellulose strips were first treated with antibody and peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) and then stained with 4-chloro-1-naphthol as described by de Wet *et al.* (24).

RESULTS

Isolation of cDNA Clones and RNA Transfer Blot Analysis. An HT-1080 cDNA library (16) containing 10^5 recombinant clones was screened with a mouse α2(IV) collagen cDNA clone (17) under cross-hybridizing conditions and one positive clone, HT-125, was identified. During subsequent rescreening of the library, three additional clones, HT-127, HT-133, and HT-141, were identified. The four overlapping positive clones contained inserts of about 540, 700, 250, and 230 nucleotides, respectively (Fig. 1), and covered one-third of the corresponding 3000-nucleotide mRNA (Fig. 2). RNA transfer blot analysis of normal human skin fibroblast and HT-1080 poly(A)$^+$ RNAs demonstrated an identical hybridization pattern (Fig. 2, lanes C and D). The clones encoded 228 amino acid residues of -Gly-Xaa-Yaa- sequences, an 18 amino acid residue-long noncollagenous, carboxyl-terminal domain, and 319 nucleotides of the 3′ untranslated region (Figs. 1 and 3).

Nucleotide and Derived Amino Acid Sequences of the cDNAs. The 5′ end of HT-125 began with a 7 amino acid residue-long noncollagenous domain and two shorter interruptions in the -Gly-Xaa-Yaa- sequence were present (Fig. 3). The 3′ end of the -Gly-Xaa-Yaa- region was relatively rich in charged amino acid residues. The carboxyl-terminal domain was only 18 amino acid residues in length. This region contained two cysteine residues but no sequences that could serve as a potential attachment site for asparagine-linked oligosaccharides (Fig. 3). The clone HT-127 contained a 229-nucleotide 3′ untranslated region and a poly(A)$^+$ tail. Two in-phase stop codons were present at the end of translation and one slightly altered poly(A) signal, AATACA, was seen. Part of HT-141 was sequenced and the nucleotide sequence was identical with HT-127 (Fig. 1). The clone HT-133 was derived from a longer mRNA than HT-127 in that it extended 90 nucleotides further 3′. HT-133 contained an additional poly(A) signal of AATAAA.

cDNA Clones HT-125 and HT-127 Show a Difference in Coding Capacity of Their Corresponding mRNAs. Comparison of nucleotide sequences of clones HT-125 and HT-127 demonstrated that HT-127 did not contain 36 nucleotides (432–467) found in HT-125 (Fig. 3). These 36 nucleotides coded for four -Gly-Xaa-Yaa- repeats and therefore did not

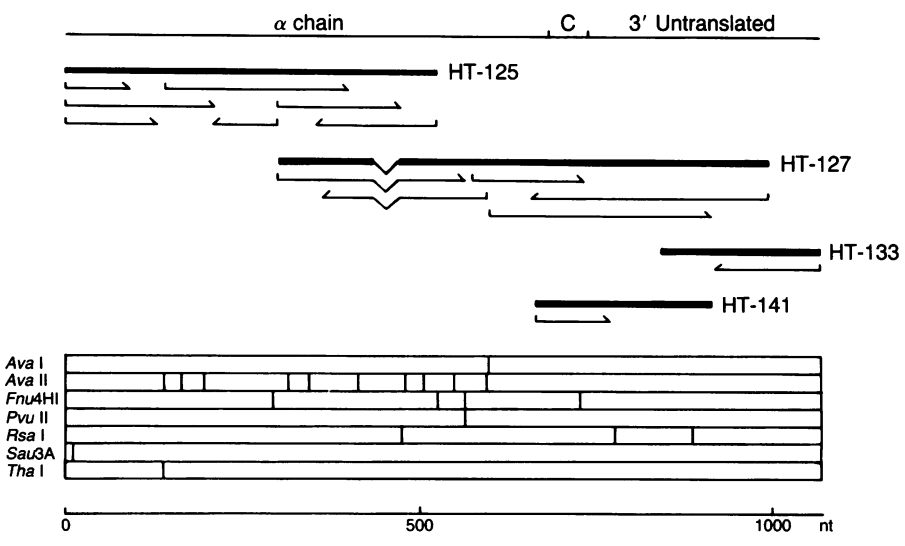

FIG. 1. Restriction map and sequencing strategies for four cloned cDNAs encoding a distinct human collagen. The strategy for sequencing is indicated with arrows representing the directions and lengths of sequencing runs. ⋁, A stretch of nucleotides not present in clone HT-127; C, the carboxyl-terminal, noncollagenous domain. Nucleotide (nt) scale is shown at bottom.

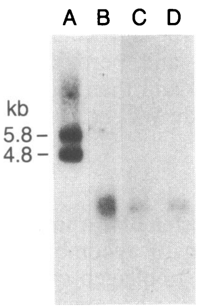

FIG. 2. RNA transfer blot analysis of poly(A)$^+$ RNAs. Lane A, 1 μg of human skin fibroblast RNA hybridized with the ^{32}P-labeled nick-translated cDNA clone α12 coding for human α1(I) collagen chain (25); lane B, 4 μg of HT-1080 RNA; lane C, 1 μg of normal human skin fibroblast RNA; and lane D, 1 μg of HT-1080 RNA hybridized with the ^{32}P-labeled, nick-translated cDNA clone HT-127. The size of the mRNA hybridizing with HT-127 was determined by using the α1(I) mRNAs of 5.8 and 4.8 kilobases (kb) as standards (26).

change the reading frame. Interestingly, the -Gly-Xaa-Yaa-repeats were uninterrupted in both clones (Fig. 3). Except for the 36-nucleotide difference, the clones HT-125 and HT-127 were identical in their overlap region that extended 133 nucleotides to the 5′ direction of the mismatch and 52 nucleotides to the 3′ direction (Figs. 1 and 3).

The difference in the two cDNAs either reflected differences in their mRNAs or was due to a cloning artifact. To determine which alternative was correct, the region of interest was analyzed by nuclease S1 mapping of the corresponding mRNAs. A 530-bp double-stranded cDNA probe was prepared (Fig. 4*A*) extending from the single *Sau*3A site (Fig. 1) to the 3′ end of HT-125. In addition to sequences complementary to mRNA, the probe contained a 15-nucleotide dC tail at the 3′ end added during cloning (16). When poly(A)$^+$ RNA from HT-1080 cells hybridized with the probe labeled at the *Sau*3A site was subjected to nuclease S1 digestion, bands of 510 and 426 nucleotides were observed (Fig. 4*B*). The 510-nucleotide band represented the fully protected cDNA probe minus the dC tail. The 426-nucleotide band was the size expected if some of the mRNA did not contain the 36 nucleotides found in HT-125 but missing in HT-127 (Fig. 4*A*). Surprisingly, a third band of 271 nucleotides was found after nuclease S1 mapping (Fig. 4*B*). The 271-nucleotide band suggested that there was another site of mismatch between the probe and the mRNA (Fig. 4 *A* and *B*). However, no overlapping clones covering this region were found and therefore the sequence of the second mismatch was not defined.

Codon Usage in the -Gly-Xaa-Yaa- Region. The nucleotides used for the third base of glycine codons (Table 1) were clearly different from those encoding the α1(I) or α2(I) chain of type I procollagen from either chicken (28) or human (27). A preference for uracil and cytosine for glycine can be seen for the type I procollagen but is not apparent in the collagen gene studied here. Instead, the codon usage for glycine showed a slight preference for adenine and guanine.

The nucleotide used for the third base of proline codons (Table 1) also showed a clear difference from chicken (28) and human type I collagen genes (27) and was similar to the pattern seen in chicken type IX (29) and human α1(IV) collagens (16).

Characterization of the Collagen by Electrophoretic Transfer Blotting. To characterize the human collagen chain(s) encoded by the mRNAs complementary to the clones reported here, antiserum was prepared in a rabbit against a synthetic peptide derived from the cDNA sequences. The synthetic peptide was 24 amino acid residues in length, and it corresponded to the last two -Gly-Xaa-Yaa- repeats and the 18 amino acid residues of the noncollagenous, carboxyl-terminal domain (Fig. 3). This stretch of amino acids was selected because it contains several charged residues and such residues usually enhance antigenicity (30). Confluent HT-1080 cells and normal human skin fibroblasts were analyzed by electrophoretic transfer blotting with antibodies to the peptide. Different dilutions of antiserum were tested and a dilution of 1:3000 was found to give the best results in these studies.

```
   1  AG  ATC CGG ACG CTG GCC TTG ATG GGG CCT CTT GGT CTT CTT GGG CAA ATT GGC CCA CTT GGA GCT CTA GGG
   1      Ile Arg Thr Leu Ala Leu Met Gly Pro Leu Gly Leu Leu Gly Gln Ile Gly Pro Leu Gly Ala Leu Gly

  72  ATG CTA GGC CAG AAG GGG GAG ATT GGA CTG CTA GGC CCT CTA GGA CAC GAT GGG AAA GGA CCT CGC GGT AAA
  24  Met Leu Gly Gln Lys Gly Glu Ile Gly Leu Leu Gly Pro Leu Gly His Asp Gly Lys Gly Pro Arg Gly Lys

 144  CTA GGA GAC ATG GGC CCT CTT GGT CCC CAA GGC CCC CTA GGA AAG GAT GGA CCT CCA GGA GTG AAG GGA GAA
  48  Leu Gly Asp Met Gly Pro Leu Gly Pro Gln Gly Pro Leu Gly Lys Asp Gly Pro Pro Gly Val Lys Gly Glu

 216  AAC GGG CAC CCA GGG AGC CCA GGA GAG AAG GGG GAA AAA GGG GAG ACA GGA CAA GCA GGC TCA CCG GTT CCT
  72  Asn Gly His Pro Gly Ser Pro Gly Glu Lys Gly Glu Lys Gly Glu Thr Gly Gln Ala Gly Ser Pro Val Pro

                          *
 288  GGG CTG CTA GGG CCA GAG GGG CCT CCC GGA CCT CCG GGG CTC CAA GGT GTT CCT GGA CCA AAG GGG GAA GCA
  96  Gly Leu Leu Gly Pro Glu Gly Pro Pro Gly Pro Pro Gly Leu Gln Gly Val Pro Gly Pro Lys Gly Glu Ala

 360  GGA CTA GAC GGA GCA AAA GGA GAG AAA GGC TTC CAG GGA GAA AAA GGA GAC CGT GGT CCC CTG GGA CTA CCC
 120  Gly Leu Asp Gly Ala Lys Gly Glu Lys Gly Phe Gln Gly Glu Lys Gly Asp Arg Gly Pro Leu Gly Leu Pro

      |-----------------------------------------------|
 432  GGA GCT TCA GGT TTG GAC GGC AGG CCT GGG CCA CCG GGT ACT CCA GGA CCA ATT GGA GTT CCA GGC CCA GCG
 144  Gly Ala Ser Gly Leu Asp Gly Arg Pro Gly Pro Pro Gly Thr Pro Gly Pro Ile Gly Val Pro Gly Pro Ala

                                  *
 504  GGA CCA AAG GGC GAG AGG GGC AGC AAA GGA GAC CCT GGG ATG ACA GGA CCA ACG GGA GCA GCT GGG CTT CCT
 168  Gly Pro Lys Gly Glu Arg Gly Ser Lys Gly Asp Pro Gly Met Thr Gly Pro Thr Gly Ala Ala Gly Leu Pro

 576  GGT TTA CAT GGA CCA CCC GGG GAC AAG GGA AAC CGG GGG GAG AGG GGG AAG AAA GGC TCT AGA GGG CCT AAA
 192  Gly Leu His Gly Pro Pro Gly Asp Lys Gly Asn Arg Gly Glu Arg Gly Lys Lys Gly Ser Arg Gly Pro Lys

                                                     ↓
 648  GGG GAC AAG GGA GAC CAA GGA GCG CCT GGA TTA GAT GCC CCC TGC CCA TTG GGG CAA GAT GGC TTA CCA GTC
 216  Gly Asp Lys Gly Asp Gln Gly Ala Pro Gly Leu Asp Ala Pro Cys Pro Leu Gly Gln Asp Gly Leu Pro Val
                              ▲

 720  CAA GGC TGC TGG AAC AAG TGA TGC CTC TAA CCT TGG ATT GGC CTG TGT GTG TGT TTG TAC ATA GAA TAT TTA
 240  Gln Gly Cys Trp Asn Lys ---         ---
                          ▲

 792  TTT TTA TAC AGT TTT CAC TTT TTG AAA ATG CCA GAA GTA TGA TGC ATC TTA CAG ATT ATT AAA AAA GAA AGA

 864  AAA ACC GTT GCA TAT TTT GTA CAG AAA ATA TCA ACC TCT TCC CTT TTG TTT ACA AGA TGT TTT GTA TAA GCC

 936  TAT GTC TCT AAT ACA TTT TTT GTT TGG TCG TAA AAA AAA AAA AAA AAA AAA AAA A
                                           AA TGT CTG CAT GAT ATT TGT GCA CAT TTA TTA AGT ATC GAA

1008  GGT TAA TAA ATT GTT GTG TCC TGG TGC CAA GGG GGG CCA GCC AGA ACT G
```

FIG. 3. Nucleotide sequences and encoded amino acid sequences of the human cDNAs. First line, nucleotide sequence of cDNAs for a distinct human collagen; second line, amino acid sequence encoded by the cDNAs. Numbering of nucleotides and amino acid residues is shown. Underlined amino acid residues indicate interruptions of the -Gly-Xaa-Yaa- sequence; the asterisk above nucleotide 302 indicates the beginning of HT-127 and the asterisk above nucleotide 522 indicates the end of HT-125. ↓, End of -Gly-Xaa-Yaa- domain; ▲, 5′ (amino acid 222) and 3′ (amino acid 245) termini of the synthetic polypeptide; ---, termination codons for translation; underlined nucleotide sequences indicate poly(A) signals. The bar (|—|) indicates the 36 nucleotides found in clone HT-125 but not in clone HT-127. Sequences starting from nucleotide 967 (second line) to 1056 are derived from clone HT-133, which extends further 3′ than HT-127. The dG tail at the beginning and the dC tail at the end of the sequences added during cloning (16) have been omitted.

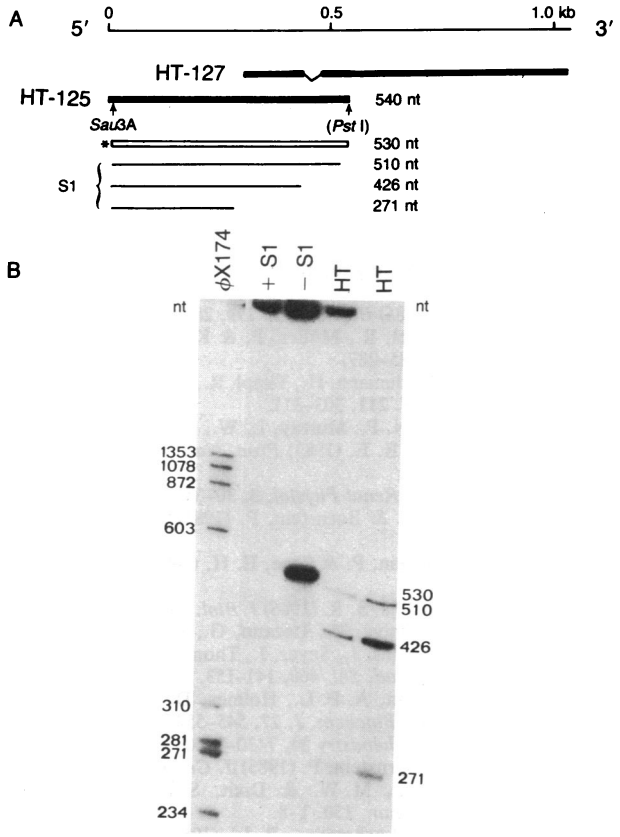

FIG. 4. Nuclease S1 mapping of poly(A)$^+$ RNA with the cDNA clone HT-125. (*A*) The probe used in the nuclease S1 mapping (open bar) is shown below cDNA clone HT-125 (black bar). A 530-nucleotide (nt)-long *Sau*3A/*Pst* I probe labeled (*) at the 3' end of the antisense strand was prepared from HT-125. The 3' *Pst* I site of HT-125 generated during cloning by dC tailing the cDNAs and dG tailing the vector pBR322 (14) is shown in parentheses. The probe covers the region (⌣) that is absent in clone HT-127 (black bar). The nuclease S1 digestion products (S1) are shown below the probe. Nucleotide scale is indicated. (*B*) Autoradiography of the nuclease S1 digestion products fractioned by gel electrophoresis. Lanes: φX174, markers derived from bacterial phage φX174 *Hae* III fragments; + S1, probe with nuclease S1 in the absence of RNA; − S1, probe without nuclease S1; HT, products of nuclease S1 digestion of hybrids of the 3' end-labeled probe and two preparations of RNA from HT-1080 cells.

Two prominent polypeptides of M_r 67,000 and M_r 62,000 were seen in the cell homogenates of HT-1080 cells (Fig. 5, lanes A and D). When using collagenous protein standards, the two major polypeptides were M_r 55,000 and M_r 43,000.

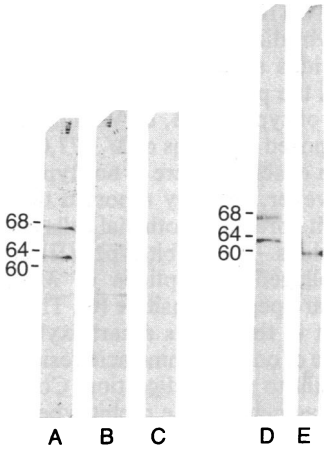

FIG. 5. Electrophoretic transfer blot analysis of HT-1080 cells. Electrophoretically fractionated HT-1080 cell (1×10^6) homogenates were transferred to nitrocellulose filters that were incubated with a 1:3000 dilution of preimmune serum (lane C) or antiserum (lanes A, B, D, and E) directed against the synthetic peptide. Lanes A, D, and E, total cell homogenate; lane B, cell homogenate digested with bacterial collagenase; lane E, cell homogenate digested with pepsin. Bovine serum albumin (M_r 68,000) and α and β subunits of prolyl-4-hydroxylase (M_r 64,000 and 60,000) were used as molecular weight markers. Molecular weights are shown as $M_r \times 10^{-3}$.

Bacterial collagenase digestion resulted in disappearance of these two polypeptides and no new ones appeared (Fig. 5, lane B). The fainter bands seen on the blots all remained after bacterial collagenase digestion. The faint bands were also detected with serum taken from the rabbit prior to immunization (Fig. 5, lane C) and thus represented reaction of the serum with some noncollagenous polypeptides. As a result of pepsin digestion, the prominent polypeptides disappeared and a new polypeptide of about M_r 60,000 appeared (Fig. 5, lane E). Also, the faint peptides disappeared as a result of pepsinization. Electrophoretic transfer blot examination of ammonium sulfate-precipitated proteins from the cell culture medium did not reveal the presence of collagenous polypeptides (not shown).

DISCUSSION

The cloned cDNAs characterized here encoded a low molecular weight human collagen that is synthesized in small amounts by cultured fibroblasts from normal skin and a human tumor cell line, HT-1080. Comparison of the nucleotide sequences presented here to nucleotide sequences for other cloned collagen chains indicated it is not α1(I) (27), α2(I) (31), α1(II) (32), α1(III) (25), α1(IV) (16), α2(IV) (17),

Table 1. Comparative codon usage in DNA sequences coding for -Gly-Xaa-Yaa- regions of several collagens

Amino acid	Human α1(I)*	Chicken α1(I)†	Chicken type IX‡	Human α1(IV)§	Low molecular weight collagen	Third base of codon
Glycine	0.50	0.57	0.41	0.29	0.11	Uracil
	0.30	0.36	0.19	0.24	0.17	Cytosine
	0.18	0.06	0.29	0.27	0.40	Adenine
	0.02	0.01	0.11	0.19	0.32	Guanine
Codons examined	261	67	198	62	72	
Proline	0.57	0.25	0.37	0.35	0.36	Uracil
	0.40	0.69	0.13	0.18	0.22	Cytosine
	0.03	0.04	0.39	0.40	0.36	Adenine
	0.00	0.02	0.11	0.07	0.06	Guanine
Codons examined	179	91	119	45	36	

*Ref. 27. †Ref. 28. ‡Ref. 29. §Ref. 16.

α2(V) (33), α1(IX) (10), α2(IX) (10), or α1(X) (34). The size of the mRNA hybridizing with the cDNA clones and the size of the polypeptides seen on electrophoretic transfer blots largely excludes the possibility that the clones encode a still uncloned chain of type V, VI, or VII collagen. If type VIII collagen is composed of chains of M_r 177,000 and M_r 125,000 (7), the collagen described here is not type VIII. Benya and Padilla (9), however, recently reported that type VIII collagen from rabbit corneal endothelial cells is composed of M_r 61,000 polypeptides. The rabbit type VIII collagen contains terminal noncollagenous peptides of M_r 14,700 and M_r 4000–5000 that are pepsin sensitive (6). The collagen studied here is different in that it has a carboxyl-terminal, noncollagenous peptide of only 18 amino acid residues (M_r 2185) that is largely resistant to pepsin digestion. Comparison of cyanogen bromide fragments of the rabbit type VIII collagen and the human collagen studied here is inconclusive since the cDNA clones cover only the last 246 amino acids. Therefore, although the collagen studied here is probably not one previously reported, it cannot safely be assigned a new roman numeral until more information is available about type VIII collagen. Also, two additional incompletely characterized low molecular weight collagens of about M_r 60,000 have been reported (13, 14). Until more information is available about them no conclusion can be reached whether they are the same or different from the low molecular weight collagen described here.

The most striking feature of the low molecular weight collagen is that its RNA transcripts undergo alternative splicing. The 36 nucleotides of coding sequences present in one of the cDNA clones (HT-125) but lacking in the other (HT-127) correspond precisely to a single 36-bp exon of the gene (L. Tikka, T.P., D.J.P., and K. Tryggvason, unpublished results). Therefore, the two cDNA clones and the two or more mRNAs detected by nuclease S1 protection experiments arose from alternative splicing of RNA transcripts of the same gene. Nuclease S1 mapping experiments demonstrated that the mRNAs complementary to both cDNA clones were present in cultured normal skin fibroblasts (not shown) as well as in HT-1080 cells. The present study provides an example of alternative splicing of a collagen gene and it will be of interest to study the biological significance of the finding.

Two bacterial collagenase-sensitive polypeptides of M_r 67,000 and M_r 62,000 were detected in cell homogenates of HT-1080 cells and cultured human skin fibroblasts (not shown) with antiserum to a synthetic peptide. The polypeptides detected with the antiserum differed in size by about 5000, or considerably more than can be accounted for by alternative splicing of 36 bp. Also, the nuclease S1 experiments suggested the presence of a second site of mismatch between the mRNAs. Therefore, the RNA transcripts may undergo alternative splicing at more than one site and the net effect of these splicing events on protein level is not yet known. It is also possible that the polypeptide chain of M_r 62,000 is a partially processed form of the M_r 67,000 chain.

The amino acid sequences derived from the cDNA clones presented several interesting features. The noncollagenous, carboxyl-terminal domain was very short, 18 amino acid residues, in contrast to most other collagens. Only in type IX collagen is the comparable domain of similar size. The collagenous sequences were interrupted by short noncollagenous domains as has been found for type IV collagen. Furthermore, the codon usage was similar to that of type IV and IX collagens. Thus, the results presented here show characterization of a low molecular weight collagen that belongs to the nonfibrillar collagens with features of type IV and type IX collagens.

We gratefully acknowledge Prof. Kari Kivirikko for valuable suggestions, Dr. Jeanne Myers for cDNA clone α12 used in this study, and Miss Aila Jokinen and Mrs. Sirkka Vilmi for expert technical assistance. This work was supported by National Institutes of Health Grant AM 16516 and by grants from the Academy of Finland and Orion Corporation Research Foundation.

1. Prockop, D. J. & Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311,** 376–386.
2. Cheah, K. S. E. (1985) *Biochem. J.* **299,** 287–303.
3. Martin, G. R., Timpl, R., Müller, P. & Kühn, K. (1985) *Trends Biochem. Sci.* **10,** 285–287.
4. Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. & Engel, J. (1983) *Biochem. J.* **211,** 303–311.
5. Bentz, H., Morris, N. P., Murray, L. W., Sakai, L. Y., Hollister, D. W. & Burgeson, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3168–3172.
6. Benya, P. D. (1980) *Renal Physiol.* **3,** 30–35.
7. Sage, H., Trüeb, B. & Bornstein, P. (1985) *J. Biol. Chem.* **258,** 13391–13401.
8. Kapoor, R., Bornstein, P. & Sage, E. H. (1986) *Biochemistry* **25,** 3930–3937.
9. Benya, P. D. & Padilla, S. R. (1986) *J. Biol. Chem.* **261,** 4160–4169.
10. Olsen, B. R., Ninomiya, Y., Lozano, G., Konomi, H., Gordon, M., Green, G., Parsons, J., Seyer, J., Thompson, H. & Vasios, G. (1985) *Ann. N.Y. Acad. Sci.* **460,** 141–153.
11. Kielty, C. M., Kwan, A. P. L., Holmes, D. F., Schor, S. L. & Grant, M. E. (1985) *Biochem. J.* **27,** 545–554.
12. Sage, H. (1985) *Biochemistry* **24,** 7430–7440.
13. Majack, R. A. & Bornstein, P. (1985) *J. Cell Biol.* **100,** 613–619.
14. Dixit, R., Harrison, M. W. & Dixit, S. N. (1985) *Biochem. Biophys. Res. Commun.* **130,** 1–8.
15. Davis, A. E. & Lachmann, P. J. (1984) *Biochemistry* **23,** 2139–2144.
16. Pihlajaniemi, T., Tryggvason, K., Myers, J. C., Kurkinen, M., Lebo, R., Cheung, M.-C., Prockop, D. J. & Boyd, D. (1985) *J. Biol. Chem.* **260,** 7681–7687.
17. Kurkinen, M., Bernard, M. P., Barlow, D. P. & Chow, L. T. (1985) *Nature (London)* **317,** 177–179.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
19. Pihlajaniemi, T. & Myers, J. C. (1987) *Methods Enzymol.*, in press.
20. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85,** 2149–2154.
21. Kivirikko, K. I. & Myllylä, R. (1987) *Methods Enzymol.*, in press.
22. de Wet, W. J., Pihlajaniemi, T., Myers, J., Kelly, T. E. & Prockop, D. J. (1983) *J. Biol. Chem.* **258,** 7721–7726.
23. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
24. de Wet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. & Helinski, D. R. (1984) *DNA* **3,** 437–447.
25. Loidl, H. R., Brinker, J. M., May, M., Pihlajaniemi, T., Morrow, S., Rosenbloom, J. & Myers, J. C. (1984) *Nucleic Acids Res.* **12,** 9383–9394.
26. Chu, M.-L., de Wet, W. J., Bernard, M. & Ramirez, F. (1985) *J. Biol. Chem.* **260,** 2315–2320.
27. Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22,** 5213–5223.
28. Tate, V., Finer, M., Boedtker, H. & Doty, P. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 1039–1049.
29. Ninomiya, Y., Showalter, A. M. & Olsen, B. R. (1984) in *The Role of Extracellular Matrix in Development*, ed. Trelstad, R. L. (Liss, New York), pp. 255–275.
30. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
31. Bernard, M. P., Myers, J. C., Chu, M.-L., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22,** 1139–1145.
32. Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E., Tsipouras, P. & Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 2207–2225.
33. Myers, J. C., Loidl, H. R., Seyer, J. M. & Dion, A. S. (1985) *J. Biol. Chem.* **260,** 11216–11222.
34. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 5041–5050.